GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2002, 22:14:33 ; Search time 56.06 Seconds
(without alignments)
462.792 Million cell updates/sec

Title: US-09-823-676-2
Perfect score: 1392
Sequence: 1 MELNRSEADEAKAETTPTGG.....IYGMSGGSGGGGGGATRPAF 270

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1392	100.0	270	2	T06118	hypothetical prote
2	706.5	50.8	311	2	F86341	hypothetical prote
3	646.5	46.4	292	2	C71448	hypothetical prote
4	645.5	46.4	302	2	H96792	unknown protein F1
5	629.5	45.2	317	2	D84890	probable AT-hook D
6	625.5	44.9	339	2	T06612	hypothetical prote
7	616	44.3	285	2	E84766	probable AT-hook D
8	605	43.5	324	2	T04576	hypothetical prote
9	581	41.7	265	2	T47898	hypothetical prote
10	572	41.1	310	2	T47695	hypothetical prote
11	554	39.8	257	2	B84860	hypothetical prote
12	417	30.0	206	2	G86279	F14L17.27 protein
13	274	19.7	296	2	G84747	AT-hook DNA-bindin
14	255.5	18.4	334	2	T04572	hypothetical prote
15	246.5	17.7	347	2	T06584	probable DNA-bindi
16	243	17.5	439	2	T52291	probable DNA-bindi
17	239.5	17.2	404	2	T05553	hypothetical prote
18	239	17.2	455	2	T05081	hypothetical prote
19	227	16.3	365	2	T06615	hypothetical prote
20	225	16.2	348	2	T02462	probable AT-hook D
21	218	15.7	345	2	T01348	hypothetical prote
22	211.5	15.2	348	2	T47923	probable DNA-bindi
23	162.5	11.7	574	2	A84782	hypothetical prote
24	146	10.5	405	2	T29167	hypothetical prote
25	139	10.0	307	2	T27609	hypothetical prote
26	137	9.8	694	2	F70868	hypothetical glyci
27	135.5	9.7	290	2	T23416	hypothetical prote
28	134.5	9.7	839	2	F75518	hypothetical prote
29	131	9.4	388	2	T29173	hypothetical prote
30	131	9.4	1381	2	E70806	hypothetical glyci
31	129.5	9.3	575	2	S35327	protein kinase sgg
32	129.5	9.3	1844	2	T51890	related to Nup98-N
33	129.5	9.3	13288	2	T03099	mucin, submaxillar
34	128.5	9.2	1585	2	T31611	hypothetical prote
35	128	9.2	2468	2	A83412	hypothetical prote
36	125.5	9.0	2038	2	A43742	female sterile hom
37	125	9.0	528	2	D46449	hypothetical prote
38	124.5	8.9	867	2	S57795	probable deoxyribo
39	124	8.9	1655	2	T13998	gene mastermind pr
40	124	8.9	3190	2	T13828	CREB-binding prote
41	122	8.8	615	2	H70589	hypothetical glyci
42	122	8.8	979	2	A35913	regulatory factor
43	121.5	8.7	670	2	T49510	fibroin-3 related
44	121.5	8.7	1231	2	S30185	insulin receptor s
45	121	8.7	418	2	B64924	hypothetical prote

ALIGNMENTS

RESULT 1
T06118
hypothetical protein F23E12.50 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 20-Jun-2000
C;Accession: T06118
R;Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Hoheis
submitted to the Protein Sequence Database, April 1999
A;Reference number: Z15485
A;Accession: T06118
A;Molecule type: DNA
A;Residues: 1-270 <BEV>
A;Cross-references: EMBL:AL022604; GSPDB:GN00062; ATSP:F23E12.50
A;Experimental source: cultivar Columbia; BAC clone F23E12
C;Genetics:
A;Gene: ATSP:F23E12.50
A;Map position: 4
C;Superfamily: Arabidopsis thaliana hypothetical protein T12H17.200

Query Match 100.0%; Score 1392; DB 2; Length 270;
Best Local Similarity 100.0%; Pred. No. 6.6e-85;
Matches 270; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MELNRSEADEAKAETTPTGGATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDSPNV 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MELNRSEADEAKAETTPTGGATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDSPNV 60

Qy      61 LRSHVLEVTSGSDISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGR 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 LRSHVLEVTSGSDISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGR 120

Qy     121 FDILSLTGTALPPPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVY 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 FDILSLTGTALPPPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVY 180

Qy     181 DRLPIEEEETPPPRTTGVQQQQPEASQSSEVTGSGAQACESNLQGGNGGGGVAFYNLGMN 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 DRLPIEEEETPPPRTTGVQQQQPEASQSSEVTGSGAQACESNLQGGNGGGGVAFYNLGMN 240

Qy     241 MNNFQFSGGDIYGMSGGSGGGGGGATRPAF 270
          ||||||||||||||||||||||||||||||
Db     241 MNNFQFSGGDIYGMSGGSGGGGGGATRPAF 270
```

RESULT 2
F86341
hypothetical protein F9H16.12 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001
C;Accession: F86341
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C. C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719
A;Accession: F86341
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-311 <STO>
A;Cross-references: GB:AE005172; NID:g4836899; PIDN:AAD30602.1; GSPDB:GN00141
C;Genetics:
A;Map position: 1
C;Superfamily: Arabidopsis thaliana hypothetical protein T12H17.200

Query Match 50.8%; Score 706.5; DB 2; Length 311;
Best Local Similarity 53.0%; Pred. No. 1.2e-39;
Matches 151; Conservative 25; Mismatches 56; Indels 53; Gaps 6;

```
Qy       7 EADEAKAETTPTGGATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDSPNVLRSHVL 66
           |:| :  :    |   |    | |: |:|||||| ||||| ||| |:|||||| ||||||
Db      59 ESDHSNKDHHQQGRPDSDPNTSSSAPGKRPRGRPPGSKNKAKPPIIVTRDSPNALRSHVL 118

Qy      67 EVTSGSDISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAP-------AGGGVITLHG 119
           ||: |:|| |:|||||| ||| || :: | | |:|||:|||  |        ||||:|||| 
Db     119 EVSPGADIVESVSTYARRRGRGVSVLGGNGTVSNVTLRQPVTPGNGGGVSGGGGVVTLHG 178

Qy     120 RFDILSLTGTALPPPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAV 179
           ||:|||||| ||||||||||||:::||||||||||||:|   |||| ||:||||||:|||
Db     179 RFEILSLTGTVLPPPAPPGAGGLSIFLAGGQGQVVGGSVVAPLIASAPVILMAASFSNAV 238

Qy     180 YDRLPIEEEETP-----------PPRTTGVQQQQPEASQSSEVTGSGAQACESNLQGGNG 228
           ::|||||||              ||:     || | ||  | ||| |          |||
Db     239 FERLPIEEEEEEGGGGGGGGGGPPQ----MQQAPSASPPSGVTGQGQ-------LGGNV 287

Qy     229 GGGVAFYNLGMNMNNFQFSGGDIYGMSGGS---GGGGGGATRPAF 270
           ||                     || ||     | | |  :|| |
Db     288 GG---------------------YGFSGDPHLLGWGAGTPSRPPF 311
```

RESULT 3
C71448
hypothetical protein - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
A;Variety: columbia
C;Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 20-Jun-2000
C;Accession: C71448
R;Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirk P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenegger, T.; Pohl, T.M.; Terryn, N.; Giel avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.
Nature 391, 485-488, 1998
A;Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech erhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Ans C.; Chalwatzis, N.
A;Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal
A;Reference number: A71400; MUID:98121113
A;Accession: C71448
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-292 <BEV>
A;Cross-references: GB:Z97344; NID:g2245126; PID:g2245139
C;Genetics:
A;Map position: 4COP9-4G3845
C;Superfamily: Arabidopsis thaliana hypothetical protein T12H17.200

Query Match 46.4%; Score 646.5; DB 2; Length 292;
Best Local Similarity 52.7%; Pred. No. 9.9e-36;
Matches 136; Conservative 32; Mismatches 47; Indels 43; Gaps 8;

```
Qy      19 GGATSSATASGSSS-----GRRPRGRPAGSKNKPKPPTIITRDSPNVLRSHVLEVTSGSD 73
           ||: ||    |        |||||||| |||||||| ||||:| | ||:|:||||:| |
Db      62 GGSGSSGGGGGHGGGGDVVGRRPRGRPPGSKNKPKPPVIITRESANTLRAHILEVTNGCD 121

Qy      74 ISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGRFDILSLTGTALPP 133
           : : |:||| ||  |:|:::||:| ||||:|||| :  | | |:|| | |:||||:|: |||
Db     122 VFDCVATYARRRQRGICVLSGSGTVTNVSIRQPS--AAGAVVTLQGTFEILSLSGSFLPP 179

Qy     134 PAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVYDRLPIEEEETPPP 193
           ||||||  ||::||||||||||||:| | | |:|||:::|||| |  |:|||:||:|
Db     180 PAPPGATSLTIFLAGGQGQVVGGSVVGELTAAGPVIVIAASFTNVAYERLPLEEDE---- 235

Qy     194 RTTGVQQQQPEASQSSEVTGSGAQACESNLQGGN--------GGGGVAFYNLGMNMN-NF 244
               ||||          | |:        |||        ||||: |:|| ||| |
Db     236 -----QQQQ---------LGGGSNG------GGNLFPEVAAGGGGGLPFFNLPMNMQPNV 275

Qy     245 QFSGGDIYGMSGGSGGGG 262
           |     : |  | ||| |
Db     276 QL---PVEGWPGNSGGRG 290
```

RESULT 4
H96792
unknown protein F14G6.10 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
C;Accession: H96792
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719
A;Accession: H96792
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-302 <STO>
A;Cross-references: GB:AE005173; NID:g6642668; PIDN:AAF20248.1; GSPDB:GN00141
C;Genetics:
A;Gene: F14G6.10
A;Map position: 1
C;Superfamily: Arabidopsis thaliana hypothetical protein T12H17.200

Query Match 46.4%; Score 645.5; DB 2; Length 302;
Best Local Similarity 49.5%; Pred. No. 1.2e-35;
Matches 138; Conservative 33; Mismatches 63; Indels 45; Gaps 7;

```
Qy       9 DEAKAETTPTGGATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDSPNVLRSHVLEV 68
           ||: :   |       |   :  |:|:|||||| |||||||| |:|||||||||||||
Db      52 DESDSNKDP-----GSDPVTSGSTGKRPRGRPPGSKNKPKPPVIVTRDSPNVLRSHVLEV 106

Qy      69 TSGSDISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAG-------GGVITLHGRF 121
           :|||:|| |:|:||| ||| || |:|| | | ||::|||| |        |||: |||||
Db     107 SSGADIVESVTTYARRRGRGVSILSGNGTVANVSLRQPATTAAHGANGGTGGVVALHGRF 166

Qy     122 DILSLTGTALPPPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVYD 181
           :|||||| |||||||:|||:::|:| ||||:|||| |:|||||:||||||:|| ::
Db     167 EILSLTGTVLPPPAPPGSGGLSIFLSGVQGQVIGGNVVAPLVASGPVILMAASFSNATFE 226

Qy     182 RLPIEEEETPPPRTTGVQQQQPEASQSSEVTGSGAQACESNLQGGNGG----------GG 231
           |||:|:|              |  :  || | |          || ||           |
Db     227 RLPLEDE------------GGEGGEGGEV-GEGG--------GGEGGPPPATSSSPPSG 264

Qy     232 VAFYNLGMNMNNFQFSGGDIYGMSGGSGGGGGGATRPAF 270
               |  ||: :    || : : |          ||||
```

```
FT   DOMAIN      480    709       LIGAND-BINDING.
FT   DOMAIN        3      6       POLY-GLN.
FT   DOMAIN      182    187       POLY-PRO.
FT   DOMAIN      201    207       POLY-ALA.
FT   DOMAIN      254    262       POLY-GLY.
SQ   SEQUENCE   709 AA;  77391 MW;  40E7666137E97B6B CRC64;


  Query Match             9.6%;  Score 133;  DB 1;  Length 709;
  Best Local Similarity   24.6%;  Pred. No. 0.15;
  Matches   70;  Conservative   32;  Mismatches  126;  Indels   56;  Gaps   11;

Qy      25 ATASGSSSG--RRPRGRPAGSKNKPKPPTIITRDSPNVLRSHV-LEVTSGSDISEAVSTY 81
           |   |||||  ||| |    ||:     | :  ||   |   | : :  | :  | :|:
Db       8 AATEGSSSGRARRPSGASTSSKDSYLGSTSVISDSAKELCKAVSVSLGLGVEALEHLSSG 67

Qy      82 ATRRG-CGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGRFDILSLTGTALPPPAPPGAG 140
             || |   : |   |   |   |     | ::      |     ||       |  |
Db      68 EQLRGDCMYAPLLGGPPVVRPTPCLPLVECKGSLL------DDGPGKGTEETAEYTPFKG 121

Qy     141 GLTVYLAGGQGQVVGGNVAGSLIASGPVVL-------------MAASFANAVYDRLPIE- 186
           |   |  | : : :| : :|   :|| : |               ||::    |   |:
Db     122 G---YNKGLEAESLGCSGSGEAGSSGTLELPSTLSLYKSGTLDEAAAYQTRDYYNFPLAL 178

Qy     187 EEETPPPRTTGVQQQQPEASQSSEVTGSGAQACESNLQGGNGGGGV-------------A 233
             : |||    :: : |     |    : ||    :|   :|||    .           :
Db     179 AGQPPPPHPRRIKLENP-LDYGSAWAAAAAQCRYGDLASLHGGGAAGPGSGSPSTAASSS 237

Qy     234 FYNLGMNMNNFQFSGGDIYGMSGGSGGGGGGA---------TRP 268
           :: |       |    | :||: || || | |           |||
Db     238 WHTL------FTTEEGQLYGLCGGGGGSGPGEAGAVAPYGYTRP 275


RESULT   2
APMU_PIG
ID   APMU_PIG       STANDARD;      PRT;  1150 AA.
AC   P12021;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Apomucin (Mucin core protein) (Fragment).
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Submaxillary gland;
RX   MEDLINE=91236743; PubMed=2033060;
RA   Eckhardt A.E., Timpte C.S., Abernethy J.L., Zhao Y., Hill R.L.;
RT   "Porcine submaxillary mucin contains a cystine-rich,
RT   carboxyl-terminal domain in addition to a highly repetitive,
RT   glycosylated domain.";
RL   J. Biol. Chem. 266:9678-9686(1991).
RN   [2]
RP   SEQUENCE OF 1-503 FROM N.A.
RC   TISSUE=Submaxillary gland;
RX   MEDLINE=88087170; PubMed=2826455;
RA   Timpte C.S., Eckhardt A.E., Abernethy J.L., Hill R.L.;
RT   "Porcine submaxillary gland apomucin contains tandemly repeated,
RT   identical sequences of 81 residues.";
RL   J. Biol. Chem. 263:1081-1088(1988).
RN   [3]
RP   SEQUENCE OF 45-80.
RC   TISSUE=Submaxillary gland;
RX   MEDLINE=87280230; PubMed=3611111;
RA   Eckhardt A.E., Timpte C.S., Abernethy J.L., Toumadje A.,
RA   Johnson W.C. Jr., Hill R.L.;
RT   "Structural properties of porcine submaxillary gland apomucin.";
RL   J. Biol. Chem. 262:11339-11344(1987).
RN   [4]
RP   CARBOHYDRATE-LINKAGE SITES, AND SEQUENCE OF 45-125.
RC   TISSUE=Submaxillary gland;
RX   MEDLINE=97248516; PubMed=9092502;
RA   Gerken T.A., Owens C.L., Pasumarthy M.;
RT   "Determination of the site-specific O-glycosylation pattern of the
RT   porcine submaxillary mucin tandem repeat glycopeptide. Model proposed
RT   for the polypeptide:galnac transferase peptide binding site.";
RL   J. Biol. Chem. 272:9709-9719(1997).
CC   -!- FUNCTION: APOMUCIN IS PART OF MUCIN, THE MAJOR GLYCOPROTEIN
CC       SYNTHESIZED AND SECRETED BY MUCOUS CELLS OF THE SUBMAXILLARY
CC       GLAND. ITS HIGHLY VISCOUS AQUEOUS SOLUTIONS SERVE TO LUBRICATE
CC       THE ORAL CAVITY AND TO PROTECT IT FROM THE EXTERNAL
CC       ENVIRONMENT.
CC   -!- SUBUNIT: INTERMOLECULAR DISULFIDE BONDS COULD HELP MAINTAIN A
CC       MULTIMERIC MUCIN STRUCTURE.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- TISSUE SPECIFICITY: SUBMAXILLARY MUCOSAE.
CC   -!- DOMAIN: CONTAINS TANDEMLY REPEATED, IDENTICAL SEQUENCES OF 81
CC       RESIDUES.
CC   -!- PTM: EXTENSIVELY O-LINKED GLYCOSYLATED ON SER AND THR RESIDUES OF
CC       THE REPEAT UNITS. HIGHEST GLYCOSYLATION APPEARS TO OCCUR ON SER
CC       RESIDUES WHICH HAVE GLY AT POSITIONS AT +2 OR -2 FROM THE
CC       GLYCOSYLATION SITE OR, WHERE GLY IS THE PENULTIMATE RESIDUE. THE
CC       PRESENCE OF PROLINE (USUALLY AT POSITION +3 OR -3) APPEARS TO ALSO
CC       ENHANCE GLYCOSYLATION.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M61883; AAA30998.1; -.
DR   EMBL; M21174; AAA30990.1; -.
DR   PIR; A40009; A40009.
DR   InterPro; IPR000359; Cys_knot.
DR   InterPro; IPR001007; VWFC.
DR   Pfam; PF00007; Cys_knot; 1.
DR   SMART; SM00041; CT; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PROSITE; PS01208; VWFC; 1.
KW   Glycoprotein; Repeat.
FT   NON_TER       1      1
FT   DOMAIN       <1    368       81 AA TANDEM REPEATS.
FT   REPEAT       <1     44       1.
FT   REPEAT       45    125       2.
FT   REPEAT      126    206       3.
FT   REPEAT      207    287       4.
FT   REPEAT      288    368       5.
FT   REPEAT      369    391       6 (INCOMPLETE).
FT   DOMAIN      929    995       VWFC.
FT   DOMAIN     1062   1146       CTCK.
FT   DISULFID   1062   1109       BY SIMILARITY.
FT   DISULFID   1076   1123       BY SIMILARITY.
FT   DISULFID   1085   1139       BY SIMILARITY.
FT   DISULFID   1089   1141       BY SIMILARITY.
FT   DISULFID      ?   1145       BY SIMILARITY.
FT   CARBOHYD     46     46       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     50     50       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     51     51       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     57     57       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     58     58       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     61     61       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     66     66       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     67     67       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     73     73       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     74     74       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     76     76       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     77     77       O-LINKED (GALNAC. . .) (MUCIN TYPE).
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2002, 22:21:23 ; Search time 49.21 Seconds
(without alignments)
212.442 Million cell updates/sec

Title: US-09-823-676-2
Perfect score: 1392
Sequence: 1 MELNRSEADEAKAETTPTGG.....IYGMSGGSGGGGGGATRPAF 270

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	133	9.6	709	1	ANDR_RABIT	P49699 oryctolagus
2	128.5	9.2	1150	1	APMU_PIG	P12021 sus scrofa
3	125.5	9.0	2038	1	FSH_DROME	P13709 drosophila
4	123.5	8.9	515	1	MEF2_DROME	P40791 drosophila
5	123.5	8.9	884	1	ANDR_EULFC	O97776 eulemur ful
6	122.5	8.8	895	1	ANDR_MACFA	O97952 macaca fasc
7	122	8.8	979	1	RFX1_HUMAN	P22670 homo sapien
8	121.5	8.7	895	1	ANDR_PAPHA	O97960 papio hamad
9	121.5	8.7	911	1	ANDR_PANTR	O97775 pan troglod
10	118.5	8.5	919	1	ANDR_HUMAN	P10275 homo sapien
11	117	8.4	367	1	BET3_MESAU	O09029 mesocricetu
12	117	8.4	1174	1	CIKE_DROME	Q02280 drosophila
13	116.5	8.4	1233	1	IRS1_MOUSE	P35569 mus musculu
14	116	8.3	1322	1	SUS_DROME	P22293 drosophila
15	115.5	8.3	378	1	CSP_PLACB	P08672 plasmodium
16	115	8.3	603	1	YD25_MYCTU	Q10637 mycobacteri
17	115	8.3	1672	1	PMPB_CHLMU	Q9pjy2 chlamydia m
18	113.5	8.2	553	1	FXC1_HUMAN	Q12948 homo sapien
19	113.5	8.2	1319	1	MN1_HUMAN	Q10571 homo sapien
20	113.5	8.2	1901	1	YZ08_MYCTU	O53553 mycobacteri
21	113.5	8.2	2090	1	N214_HUMAN	P35658 homo sapien
22	113	8.1	801	1	Y747_MYCTU	O53810 mycobacteri
23	113	8.1	2333	1	PGCA_CANFA	Q28343 canis famil
24	112.5	8.1	1235	1	IRS1_RAT	P35570 rattus norv
25	112	8.0	543	1	YP91_MYCTU	Q50630 mycobacteri
26	110.5	7.9	654	1	TFE2_HUMAN	P15923 homo sapien
27	110.5	7.9	860	1	ELS_MOUSE	P54320 mus musculu
28	110.5	7.9	1156	1	GLH4_CAEEL	O76743 caenorhabdi
29	110.5	7.9	2517	1	NCR2_HUMAN	Q9y618 h nuclear r
30	109.5	7.9	589	1	DYRB_MOUSE	Q9z188 mus musculu
31	109	7.8	672	1	PHX5_MOUSE	P08399 mus musculu
32	109	7.8	732	1	TAU_MOUSE	P10637 mus musculu
33	109	7.8	825	1	ICP0_HSV2H	P28284 herpes simp
34	109	7.8	910	1	ILF3_RAT	Q9jil3 rattus norv
35	108.5	7.8	433	1	HXB3_MOUSE	P09026 mus musculu
36	108.5	7.8	864	1	ELS_RAT	Q99372 rattus norv
37	108.5	7.8	914	1	WA22_MYCTU	O06794 mycobacteri
38	108	7.8	323	1	JUND_CHICK	P27921 gallus gall
39	108	7.8	1596	1	MAM_DROME	P21519 drosophila
40	107.5	7.7	843	1	MNB_DROME	P49657 drosophila
41	107	7.7	757	1	TAU_HUMAN	P10636 homo sapien
42	107	7.7	2318	1	NTC3_MOUSE	Q61982 mus musculu
43	106.5	7.7	316	1	LORI_HUMAN	P23490 homo sapien
44	106.5	7.7	332	1	HME2_HUMAN	P19622 homo sapien
45	106.5	7.7	660	1	YHL1_EBV	P03181 epstein-bar

ALIGNMENTS

```
RESULT   1
ANDR_RABIT
ID   ANDR_RABIT     STANDARD;      PRT;   709 AA.
AC   P49699;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Androgen receptor (Dihydrotestosterone receptor) (Fragment).
GN   AR OR NR3C4.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NEW ZEALAND WHITE; TISSUE=Prostate;
RX   MEDLINE=96044663; PubMed=7559153;
RA   Krongrad A., Wilson J.D., McPhaul M.J.;
RT   "Cloning and partial sequence of the rabbit androgen receptor:
RT   expression in fetal urogenital tissues.";
RL   J. Androl. 16:209-212(1995).
CC   -!- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN
CC       THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR
CC       PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC       A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
CC       NR3 SUBFAMILY.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; U16366; AAC48469.1; -.
DR   HSSP; P06536; 1GDC.
DR   InterPro; IPR001103; Androgen_recep.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001628; zf-C4.
DR   Pfam; PF02166; Androgen_recep; 1.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW   Zinc-finger; Steroid-binding.
FT   NON_TER       1      1
FT   DOMAIN       <1    347       MODULATING.
FT   DNA_BIND     349    414       NUCLEAR RECEPTOR-TYPE.
FT   ZN_FING      349    369       C4-TYPE.
FT   ZN_FING      385    409       C4-TYPE.
```

```
Qy      84 RRGCGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGRFDILSLTGTALPPPAP---PGAG 140
           :    :|::   | |::||:|||   : || :|  |||:||||:|| :|  :       |
Db     173 QGSLAICVLCANGVVSSVTLRQP--DSSGGTLTYEGRFEILSLSGTFMPSDSDGTRSRTG 230

Qy     141 GLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVYDRLPIEEEETPPP 193
           |::| ||   |:|||| ||| |:|: |: ::  :|          ::|:|| |
Db     231 GMSVSLASPDGRVVGGGVAGLLVAATPIQVVVGTFLGGTN-----QQEQTPKP 278


RESULT  15
T06584
probable DNA-binding protein - garden pea
C;Species: Pisum sativum (garden pea)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 01-Dec-2000
C;Accession: T06584; T06582
R;Sato, N.; Kazuno, A.A.; Ohta, N.; Ohshima, K.
submitted to the EMBL Data Library, June 1996
A;Description: Identification of a novel family of DNA-binding proteins with two AT-hook
A;Reference number: Z15774
A;Accession: T06584
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-347 <SAT>
A;Cross-references: EMBL:X98738; PIDN:CAA67290.1
A;Accession: T06582
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-114,'S',116-334 <SA2>
A;Cross-references: EMBL:X98739; PIDN:CAA67291.1
A;Experimental source: cv. Alaska
C;Keywords: DNA binding


  Query Match             17.7%;  Score 246.5;  DB 2;  Length 347;
  Best Local Similarity   36.7%;  Pred. No. 2.8e-09;
  Matches   61;  Conservative   30;  Mismatches   66;  Indels    9;  Gaps    5;

Qy      16 TPTGGATSSATASGSSSGRRPRGRPAGSKNKPKPPTI---ITRDSPNVLRSHVLEVTSGS 72
           ||     :|   ||: | :| |||| ||  | :   :  :| :       ||: : :|
Db     115 TPMSAPANSTQDSGTPSEKRGRGRPRGSGRKQQLAALGDWMTSSAGLAFSPHVITIAAGE 174

Qy      73 DISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGRFDILSLTGTAL- 131
           ||:  :   : :|   :||:||||  : ||:||||:   |  :|  |:| ||||:|: |
Db     175 DIAAKLLLLSQQRPRALCILSGTGIASKVTLRQPASTNAG--VTYEGKFQILSLSGSYLV 232

Qy     132 -PPPAPPG-AGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASF 175
                |    ||::| |:   | |:||:||  |||  |: |:  ||
Db     233 SEDGGPTNRTGGISVSLSSRDGHVIGGSVA-MLIAGSPIQLVVCSF 277


Search completed: July 10, 2002, 22:26:43
Job time: 730 sec
```

```
  Query Match             39.8%;  Score 554;  DB 2;  Length 257;
  Best Local Similarity   52.8%;  Pred. No. 1.1e-29;
  Matches  113;  Conservative   30;  Mismatches   37;  Indels   34;  Gaps    4;

Qy      16 TPTGGATSS-------ATASGSSS-------------------GRRPRGRPAGSKNKPK 48
           |||   : |       ||  |:::                   :|||||||||||||||
Db       8 TPTSVGSKSVPMRNHEATERGNTNNNLRALPKAVQPVSSIEGEMAKRPRGRPAGSKNKPK 67

Qy      49 PPTIITRDSPNVLRSHVLEVTSGSDISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAA 108
           || |:| |||| ||:: :|::|| || | :| :| |:  |:||:|  | |||||:|||
Db      68 PPIIVTHDSPNSLRANAVEISSGCDICETLSDFARRKQRGLCILSANGCVTNVTLRQPA- 126

Qy     109 PAGGGVITLHGRFDILSLTGTALPPPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPV 168
           : | ::|||||::|||| |: |||||| |  |||:|||| ||||||| | | |||||||
Db     127 -SSGAIVTLHGRYEILSLLGSILPPPAPLGITGLTIYLAGPQGQVVGGGVVGGLIASGPV 185

Qy     169 VLMAASFANAVYDRLPIEEEETPPPRTTGVQQQQ 202
           ||||||| |||:||||::::|        :| ||
Db     186 VLMAASFMNAVFDRLPMDDDE-----AASMQNQQ 214


RESULT  12
G86279
F14L17.27 protein - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Nov-2001
C;Accession: G86279
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719
A;Accession: G86279
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-206 <STO>
A;Cross-references: GB:AE005172; NID:g7262692; PIDN:AAF43950.1; GSPDB:GN00141
C;Genetics:
A;Map position: 1


  Query Match             30.0%;  Score 417;  DB 2;  Length 206;
  Best Local Similarity   46.0%;  Pred. No. 9.1e-21;
  Matches   91;  Conservative   32;  Mismatches   59;  Indels   16;  Gaps    5;

Qy      35 RPRGRPAGSKNKPKPPTIITRDSPNVLRSHVLEVTSGSDISEAVSTYATRRGCGVCIISG 94
           |||||| ||||||| |  :| | |  :  ::||| ||:|: ||:: :   :  | |::||
Db       6 RPRGRPRGSKNKPKAPIFVTIDPP--MSPYILEVPSGNDVVEALNRFCRGKAIGFCVLSG 63

Qy      95 TGAVTNVTIRQPAAPAGGGVITLHGRFDILSLTGTALPP-----PAPPGAGGLTVYLAGG 149
           :|:| :||:|||:  | |  || ||:||:||:: | |||      :|| :   || |||
Db      64 SGSVADVTLRQPSPAAPGSTITFHGKFDLLSVSATFLPPLPPTSLSPPVSNFFTVSLAGP 123

Qy     150 QGQVVGGNVAGSLIASGPVVLMAASFANAVYDRLPIEEEETPPPRTTGVQQQQPEASQSS 209
           ||:|:|| ||| |:|:| |  :| || |  | |||  |||    | :   :  |  ||
Db     124 QGKVIGGFVAGPLVAAGTVYFVATSFKNPSYHRLPATEEE---QRNSA---EGEEEGQSP 177

Qy     210 EVTGSGAQACESNLQGGN 227
           |:| |    ||   ||:
Db     178 PVSGGGG---ESMYVGGS 192


RESULT  13
G84747
AT-hook DNA-binding protein (AHP1) [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: G84747
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487
A;Accession: G84747
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-296 <STO>
A;Cross-references: GB:AE002093; NID:g2459442; PIDN:AAB80677.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g33620
A;Map position: 2


  Query Match             19.7%;  Score 274;  DB 2;  Length 296;
  Best Local Similarity   35.1%;  Pred. No. 3.6e-11;
  Matches   87;  Conservative   28;  Mismatches   81;  Indels   52;  Gaps   11;

Qy      22 TSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDSPNVLRS-------HVLEVTSGSDI 74
           | |  :||   | : |||| || :|        |     | |       ||| | :| |:
Db      70 TVSQPSSGGDGGEKKRGRPPGSSSK--------RLKLQALGSTGIGFTPHVLTVLAGEDV 121

Qy      75 SEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGRFDILSLTGT--ALP 132
           |  :          ||::|  ||::|||:|| |    || :|  |||:||||:|:   |
Db     122 SSKIMALTHNGPRAVCVLSANGAISNVTLRQSA--TSGGTVTYEGRFEILSLSGSFHLLE 179

Qy     133 PPAPPG-AGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVYDRLPIEEEE-- 189
                 |||:| |:   | |:||:||| |||: || ::  ||       ||  |:|
Db     180 NNGQRSRTGGLSVSLSSPDGNVLGGSVAGLLIAASPVQIVVGSF-------LPDGEKEPK 232

Qy     190 --------TPP--PRTTGVQQQQPEASQSSEVTGSGAQACESNLQGGNGG------GGVA 233
                  : |  ||     |     :|  |  | |     ||:  ||:|       ||
Db     233 QHVGQMGLSSPVLPRVAPTQVLMTPSSPQSRGTMS-----ESSCGGGHGSPIHQSTGGP- 286

Qy     234 FYNLGMNM 241
           ||  :||
Db     287 -YNNTINM 293


RESULT  14
T04572
hypothetical protein T12H17.160 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 14-May-1999
C;Accession: T04572
R;Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancro
submitted to the Protein Sequence Database, February 1998
A;Reference number: Z15377
A;Accession: T04572
A;Molecule type: DNA
A;Residues: 1-334 <BEV>
A;Cross-references: EMBL:AL021635
A;Experimental source: cultivar Columbia; BAC clone T12H17
C;Genetics:
A;Map position: 4
A;Introns: 139/1; 161/3; 205/3
A;Note: T12H17.160


  Query Match             18.4%;  Score 255.5;  DB 2;  Length 334;
  Best Local Similarity   27.5%;  Pred. No. 6.8e-10;
  Matches   64;  Conservative   38;  Mismatches   68;  Indels   63;  Gaps    6;

Qy      17 PTGGATSSATASGSSSG--RRPRGRP---------------------------------- 40
           |    | ||   | |||  :: ||||
Db      53 PQNSFTPSAAMDGFSSGPIKKRRGRPRKYGHDGAAVTLSPNPISSAAPTTSHVIDFSTTS 112

Qy      41 -AGSKNKPKPPTIITRDSPNV---------------LRSHVLEVTSGSDISEAVSTYAT 83
            | ||  ||  :  |                    |:: | :| |::: : :::
Db     113 EKRGKMKPATPTPSSFIRPKYQVENLGEWSPSSAAANFTPHIITVNAGEDVTKRIISFSQ 172
```

R;Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft,
submitted to the Protein Sequence Database, February 1998
A;Reference number: Z15377
A;Accession: T04576
A;Molecule type: DNA
A;Residues: 1-324 <BEV>
A;Cross-references: EMBL:AL021635
A;Experimental source: cultivar Columbia; BAC clone T12H17
C;Genetics:
A;Map position: 4
A;Note: T12H17.200
C;Superfamily: Arabidopsis thaliana hypothetical protein T12H17.200

Query Match 43.5%; Score 605; DB 2; Length 324;
Best Local Similarity 56.8%; Pred. No. 6e-33;
Matches 129; Conservative 26; Mismatches 42; Indels 30; Gaps 6;

```
Qy      19 GGATSSATASGSSSG-----RRPRGRPAGSKNKPKPPTIITRDSPNVLRSHVLEVTSGSD 73
           ||   |    | | |     |||||||||||||||| |||||| | ||:||:|:  | |
Db      88 GG---SGEGGGGSGGDHQMTRRPRGRPAGSKNKPKPPIIITRDSANALRTHVMEIGDGCD 144

Qy      74 ISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAA-PAGGGVITLHGRFDILSLTGTALP 132
           : |:|:|:| ||  |||::|||| ||||||||| : |: | |::|||||:||||:|: ||
Db     145 LVESVATFARRRQRGVCVMSGTGNVTNVTIRQPGSHPSPGSVVSLHGRFEILSLSGSFLP 204

Qy     133 PPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVYDRLPIEEEE--T 190
           ||||| | ||:|||||||||||||:| | |: :||||:|||||:|| |:|||:||:|  |
Db     205 PPAPPTATGLSVYLAGGQGQVVGGSVVGPLLCAGPVVVMAASFSNAAYERLPLEEDEMQT 264

Qy     191 P-----------PPRTTGVQQQQPEASQSSEVTGSGAQACESNLQG 225
           |           ||      | | :|        || |    || |
Db     265 PVHGGGGGGSLESPPMMGQQLQHQQQAM-------SGHQGLPPNLLG 304
```

RESULT 9
T47898
hypothetical protein T4C21.280 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 02-Sep-2000
C;Accession: T47898
R;Choisne, N.; Robert, C.; Brottier, P.; Wincker, P.; Cattolico, L.; Artiguenave, F.; Sa
.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.
submitted to the Protein Sequence Database, March 2000
A;Reference number: Z24479
A;Accession: T47898
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-265 <CHO>
A;Cross-references: EMBL:AL162295
A;Experimental source: cultivar Columbia; BAC clone T4C21
C;Genetics:
A;Map position: 3
A;Note: T4C21.280
C;Superfamily: Arabidopsis thaliana hypothetical protein T12H17.200

Query Match 41.7%; Score 581; DB 2; Length 265;
Best Local Similarity 50.2%; Pred. No. 1.9e-31;
Matches 119; Conservative 31; Mismatches 57; Indels 30; Gaps 3;

```
Qy      34 RRPRGRPAGSKNKPKPPTIITRDSPNVLRSHVLEVTSGSDISEAVSTYATRRGCGVCIIS 93
           |||||||||||||| | |:|||| |  | ||:|:|:  |: |::: :| ||  |||::
Db      59 RRPRGRPAGSKNKPKAPIIVTRDSANAFRCHVMEITNACDVMESLAVFARRRQRGVCVLT 118

Qy      94 GTGAVTNVTIRQPAAPAGGGVITLHGRFDILSLTGTALPPPAPPGAGGLTVYLAGGQGQV 153
           | |||||||:|||    ||||::|||||:||||:|: ||||||| | || ||||||||||
Db     119 GNGAVTNVTVRQP----GGGVVSLHGRFEILSLSGSFLPPPAPPAASGLKVYLAGGQGQV 174

Qy     154 VGGNVAGSLIASGPVVLMAASFANAVYDRLPIEEEETPPPRTTGVQQQQPEASQSSEVTG 213
           :|||:| | | || |||:||||| || |:|||:|||| |: |
Db     175 IGGSVVGPLTASSPVVVMAASFGNASYERLPLEEEE----------------ETEREIDG 218

Qy     214 SGAQACESNLQGGNGGGGVAFYNLGMNMNNFQFSGGDIYGMSGGSGGGGGGATRPAF 270
           : |:|  :  |         :|     |: |     |: |              |  ||:|
Db     219 NAARAIGTQTQKQLMQDATSFIGSPSNLINSVSLPGEAY----------WGTQRPSF 265
```

RESULT 10
T47695
hypothetical protein T22E16.220 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Jun-2000
C;Accession: T47695
R;Benes, V.; Wurmbach, E.; Drzonek, H.; Ansorge, W.; Mewes, H.W.; Lemcke, K.; Mayer,
submitted to the Protein Sequence Database, February 2000
A;Reference number: Z24472
A;Accession: T47695
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-310 <BEN>
A;Cross-references: EMBL:AL132975
A;Experimental source: cultivar Columbia; BAC clone T22E16
C;Genetics:
A;Map position: 3
A;Note: T22E16.220
C;Superfamily: Arabidopsis thaliana hypothetical protein T12H17.200

Query Match 41.1%; Score 572; DB 2; Length 310;
Best Local Similarity 47.8%; Pred. No. 8.6e-31;
Matches 120; Conservative 44; Mismatches 69; Indels 18; Gaps 6;

```
Qy       5 RSEADEAKAETTPTGGATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDSPNVLRSH 64
           :|: ::   :  |   |    : || |:|||||||| |||||||| | ::|::||| |:||
Db      63 QSQEEQNSRDEQP---AVEPGSGSG-STGRRPRGRPPGSKNKPKSPVVVTKESPNSLQSH 118

Qy      65 VLEVTSGSDISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGRFDIL 124
           |||: :|:|::|::: :| ||| || ::||:| |||||:|||  | |||::| |:|:||
Db     119 VLEIATGADVAESLNAFARRRGRGVSVLSGSGLVTNVTLRQPA--ASGGVVSLRGQFEIL 176

Qy     125 SLTGTALPPP-APPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVYDRL 183
           |: |  ||   :|  | |||:|||| ||||||| ||| |||||||:::||:| || |:||
Db     177 SMCGAFLPTSGSPAAAAGLTIYLAGAQGQVVGGGVAGPLIASGPVIVIAATFCNATYERL 236

Qy     184 PIEEEETPPPRTTGVQQQQPEASQSSEVTGSGAQACESNLQGGNGGGGVAFYNLGMNM-- 241
           |||||:         ||:||   :  :          ||   |  |      ||:  |
Db     237 PIEEEQ---------QQEQPLQLEDGKKQKEENDDNESGNNGNEGSMQPPMYNMPPNFIP 287

Qy     242 NNFQFSGGDIY 252
           |  | :  |:|
Db     288 NGHQMAQHDVY 298
```

RESULT 11
B84860
hypothetical protein At2g42940 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: B84860
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487
A;Accession: B84860
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-257 <STO>
A;Cross-references: GB:AE002093; NID:g4512661; PIDN:AAD21715.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g42940
A;Map position: 2

```
Db     265 AGQGQLRGNMSGYDQFAGDPH-LLGWGAAAAAAPPRPAF 302


RESULT   5
D84890
probable AT-hook DNA-binding protein [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C;Accession: D84890
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487
A;Accession: D84890
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-317 <STO>
A;Cross-references: GB:AE002093; NID:g2583112; PIDN:AAB82621.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g45430
A;Map position: 2
C;Superfamily: Arabidopsis thaliana hypothetical protein T12H17.200


  Query Match             45.2%;  Score 629.5;  DB 2;  Length 317;
  Best Local Similarity   54.1%;  Pred. No. 1.4e-34;
  Matches  126;  Conservative   37;  Mismatches   47;  Indels   23;  Gaps    5;

Qy       2 ELNRSEADEA-KAETTPTGGATSSATASGSSS-GRRPRGRPAGSKNKPKPPTIITRDSPN 59
           ::: :|   | | ::||  |  |     | :   |||||||||||||||||| |||||| |
Db      55 DIDPNEHSSAGKDQSTPGSGGESGGGGGGDNHITRRPRGRPAGSKNKPKPPIIITRDSAN 114

Qy      60 VLRSHVLEVTSGSDISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGG--VITL 117
           |:|||:|| :| |: |:|: :| ||  |:|:::|| ||||||||||||:  |||  |: |
Db     115 ALKSHVMEVANGCDVMESVTVFARRRQRGICVLSGNGAVTNVTIRQPASVPGGGSSVVNL 174

Qy     118 HGRFDILSLTGTALPPPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFAN 177
           ||||:||||:|: ||||||| | |||:||||||||||||:| | |:|||||:||||| |
Db     175 HGRFEILSLSGSFLPPPAPPAASGLTIYLAGGQGQVVGGSVVGPLMASGPVVIMAASFGN 234

Qy     178 AVYDRLPIEEEETPPPRTTGVQQQQPEASQSSEVTGSGAQACESNLQGGNGGG 230
           | |:|||:||::         |::|   : :: : |:          |||
Db     235 AAYERLPLEEDD---------QEEQTAGAVANNIDGNATM----------GGG 268


RESULT   6
T06612
hypothetical protein F16J13.120 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 20-Jun-2000
C;Accession: T06612
R;Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft,
submitted to the Protein Sequence Database, April 1999
A;Reference number: Z15789
A;Accession: T06612
A;Molecule type: DNA
A;Residues: 1-339 <BEV>
A;Cross-references: EMBL:AL049638; GSPDB:GN00062; ATSP:F16J13.120
A;Experimental source: cultivar Columbia; BAC clone F16J13
C;Genetics:
A;Gene: ATSP:F16J13.120
A;Map position: 4
C;Superfamily: Arabidopsis thaliana hypothetical protein T12H17.200


  Query Match             44.9%;  Score 625.5;  DB 2;  Length 339;
  Best Local Similarity   49.2%;  Pred. No. 2.8e-34;
  Matches  129;  Conservative   28;  Mismatches   48;  Indels   57;  Gaps    3;

Qy       4 NRSEADEAKAETTPTGGATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDSPNVLRS 63
           |,:  | |  :  |  |    ||   |||||||||||||| | |||||| | ||:

Db      88 NTNSGSEGKEMSLHGGEGGSGGGGSGEQMTRRPRGRPAGSKNKPKAPIIITRDSANALRT 147

Qy      64 HVLEVTSGSDISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGRFDI 123
           ||:|:  | || : ::|:| ||  |||::||||:||||||||| :|  | |::|||||:|
Db     148 HVMEIGDGCDIVDCMATFARRRQRGVCVMSGTGSVTNVTIRQPGSPP-GSVVSLHGRFEI 206

Qy     124 LSLTGTALPPPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVYDRL 183
           |||:|: |||||| | ||:|||||||||||:| | |: |||||:|||||:|| |:||
Db     207 LSLSGSFLPPPAPPAATGLSVYLAGGQGQVVGGSVVGPLLCSGPVVVMAASFSNAAYERL 266

Qy     184 PIEEEETPPPRTTGVQQQQPEASQSSEVTGSGAQACESNLQGGNGGGGVAFYNLGMNMNN 243
           |:||:|   |                                :||| ||||
Db     267 PLEEDEMQTP----------------------------VQGGGGGGG------------ 285

Qy     244 FQFSGGDIYGMSGGSGGGGGGA 265
                       |||| |:
Db     286 ---------------GGGGMGS 292


RESULT   7
E84766
probable AT-hook DNA-binding protein [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C;Accession: E84766
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487
A;Accession: E84766
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-285 <STO>
A;Cross-references: GB:AE002093; NID:g3668079; PIDN:AAC61811.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g35270
A;Map position: 2
C;Superfamily: Arabidopsis thaliana hypothetical protein T12H17.200


  Query Match             44.3%;  Score 616;  DB 2;  Length 285;
  Best Local Similarity   49.6%;  Pred. No. 9.9e-34;
  Matches  130;  Conservative   30;  Mismatches   64;  Indels   38;  Gaps    6;

Qy      11 AKAETTPTGGATSSATASGSSSG----RRPRGRPAGSKNKPKPPTIITRDSPNVLRSHVL 66
           |  : :  ||        |  ||    |||||||||||||||| |:||:| | ||:|:|
Db      57 ASNDNSGLGG------GGGGGSGDLVMRRPRGRPAGSKNKPKPPVIVTRESANTLRAHIL 110

Qy      67 EVTSGSDISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGRFDILSL 126
           || || |: | :|||| ||  |:|::|||| ||||:||||   | | |:|| | |:|||
Db     111 EVGSGCDVFECISTYARRRQRGICVLSGTGTVTNVSIRQPT--AAGAVVTLRGTFEILSL 168

Qy     127 TGTALPPPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVYDRLPIE 186
           :|: |||||||||  ||::||| |||||||| | |:|:|||::|||| |  |:|||::
Db     169 SGSFLPPPAPPGATSLTIFLAGAQGQVVGGNVVGELMAAGPVMVMAASFTNVAYERLPLD 228

Qy     187 EEETPPPRTTGVQQQQPEASQSSEVTGSGAQACESNLQGGNGGGGVAFYNLGMNMNNFQF 246
           | |               |  ||    | |    |:      ||||: |:|| |:|
Db     229 EHE--------------EHLQSGGGGGGGNMYSEAT----GGGGGLPFFNLPMSMPQI-- 268

Qy     247 SGGDIYGMSGGSGGGGGGATRP 268
                 |:   |  |   |
Db     269 ------GVESWQGNHAGAGRAP 284


RESULT   8
T04576
hypothetical protein T12H17.200 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 09-Jun-2000
C;Accession: T04576
```

```
RESULT  14
US-09-219-849-50
; Sequence 50, Application US/09219849
; Patent No. 6150081
; GENERAL INFORMATION:
;  APPLICANT:  VAN HEERDE, GEORGE V.
;  APPLICANT:  VAN RIJN, ALEXIS C.
;  APPLICANT:  BOUWSTRA, JAN B.
;  APPLICANT:  DE WOLF, FREDERIK A.
;  APPLICANT:  MOOBROEK, ANDREAS
;  APPLICANT:  WERTEN, MARC W.T.
;  APPLICANT:  WIND, RICHELE D.
;  APPLICANT:  VAN DEN BOSCH, TANJA J.
;  TITLE OF INVENTION:  SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
;  TITLE OF INVENTION:  SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
;  TITLE OF INVENTION:  PREPARATION THEREOF
;  FILE REFERENCE:  2728-2
;  CURRENT APPLICATION NUMBER: US/09/219,849
;  CURRENT FILING DATE:  1998-12-23
;  NUMBER OF SEQ ID NOS:  50
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 50
;   LENGTH: 595
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
;   OTHER INFORMATION: amino acid sequence
US-09-219-849-50


  Query Match             7.3%;  Score 101.5;  DB 4;  Length 595;
  Best Local Similarity   23.4%;  Pred. No. 0.74;
  Matches   71;  Conservative   19;  Mismatches  118;  Indels   95;  Gaps   15;

Qy       6 SEADEAKAETTPTG--GATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDSPNVLRS 63
           |  | || :  | |  |   |   :|:     ||| | | : :| ||
Db      92 SGLDGAKGDAGPAGPKGEPGSPGENGAPGQMGPRGLP-GERGRPGPP------------ 137

Qy      64 HVLEVTSGSDISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAG-GGVITLHGRFD 122
                     |   |   |    :|    |       | | |  | :   |
Db     138 --------------GTAGARGNDGAVGAAGPPGPTG-----PTGPPGFPGAVGAKGEAG 177

Qy     123 ILSLTGTALP-----PPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASG-PVVLMAASFA 176
             |:  |       | |||  |     ||  |        |:  |:| | :  |  |
Db     178 PQGARGSEGPQGVRGEPGPPGPAG----AAGPAGNPGADGQPGAKGANGAPGIAGAPGFP 233

Qy     177 NAVYDRLPIEEEETPP----------------------PRTTGVQQQQPEASQSSEVTG 213
           |     | :   ||                        |  ||||     | :  :  |
Db     234 GARGPSGP-QGPSGPPGPKGNSGEPGAPGNKGDTGAKGEPGATGVQGPPGPAGEEGK-RG 291

Qy     214 SGAQACESNL------QGGNGGGGVAFYNLGMNMNNFQFSGGDIYGMSGGSG-GGGGGAT 266
           :  :   | |      :|| |  |              | | |  |::|  |  |  ||
Db     292 ARGEPGPSGLPGPPGERGGPGSRG-------------FPGAD--GVAGPKGPSGERGAP 335

Qy     267 RPA 269
             ||
Db     336 GPA 338


RESULT  15
US-09-219-849-49
; Sequence 49, Application US/09219849
; Patent No. 6150081
; GENERAL INFORMATION:
;  APPLICANT:  VAN HEERDE, GEORGE V.
;  APPLICANT:  VAN RIJN, ALEXIS C.
;  APPLICANT:  BOUWSTRA, JAN B.
;  APPLICANT:  DE WOLF, FREDERIK A.
;  APPLICANT:  MOOBROEK, ANDREAS
;  APPLICANT:  WERTEN, MARC W.T.
;  APPLICANT:  WIND, RICHELE D.
;  APPLICANT:  VAN DEN BOSCH, TANJA J.
;  TITLE OF INVENTION:  SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
;  TITLE OF INVENTION:  SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
;  TITLE OF INVENTION:  PREPARATION THEREOF
;  FILE REFERENCE:  2728-2
;  CURRENT APPLICATION NUMBER: US/09/219,849
;  CURRENT FILING DATE:  1998-12-23
;  NUMBER OF SEQ ID NOS:  50
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 49
;   LENGTH: 822
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
;   OTHER INFORMATION: amino acid sequence
US-09-219-849-49


  Query Match             7.3%;  Score 101.5;  DB 4;  Length 822;
  Best Local Similarity   23.4%;  Pred. No. 1.1;
  Matches   71;  Conservative   19;  Mismatches  118;  Indels   95;  Gaps   15;

Qy       6 SEADEAKAETTPTG--GATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDSPNVLRS 63
           |  | || :  | |  |   |   :|:     ||| | | : :| ||
Db      92 SGLDGAKGDAGPAGPKGEPGSPGENGAPGQMGPRGLP-GERGRPGPP------------ 137

Qy      64 HVLEVTSGSDISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAG-GGVITLHGRFD 122
                     |   |   |    :|    |       | | |  | :   |
Db     138 --------------GTAGARGNDGAVGAAGPPGPTG-----PTGPPGFPGAVGAKGEAG 177

Qy     123 ILSLTGTALP-----PPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASG-PVVLMAASFA 176
             |:  |       | |||  |     ||  |        |:  |:| | :  |  |
Db     178 PQGARGSEGPQGVRGEPGPPGPAG----AAGPAGNPGADGQPGAKGANGAPGIAGAPGFP 233

Qy     177 NAVYDRLPIEEEETPP----------------------PRTTGVQQQQPEASQSSEVTG 213
           |     | :   ||                        |  ||||     | :  :  |
Db     234 GARGPSGP-QGPSGPPGPKGNSGEPGAPGNKGDTGAKGEPGATGVQGPPGPAGEEGK-RG 291

Qy     214 SGAQACESNL------QGGNGGGGVAFYNLGMNMNNFQFSGGDIYGMSGGSG-GGGGGAT 266
           :  :   | |      :|| |  |              | | |  |::|  |  |  ||
Db     292 ARGEPGPSGLPGPPGERGGPGSRG-------------FPGAD--GVAGPKGPSGERGAP 335

Qy     267 RPA 269
             ||
Db     336 GPA 338


Search completed: July 10, 2002, 22:21:18
Job time: 2810 sec
```

```
Db   1425 AAHGHGAHRVRQGPERVLGGHGVPDVRQRRGHAAGDEGAVAV-GRVDPALAELVEALVGG 1483

Qy    130 ALPPPAPPG---AGGLTVYLA-------------------GGQGQVVGGNVAGSLIASG 166
Db   1484 LHPRVLQPGHGLAGGPAVDEAHEGLVLLPRVPHLRDEDGHGPGRGAVAGRGLADVVLVPE 1543

Qy    167 PVV-LMAASFANAVYDR-------LPIEEEETPPPRTTGVQQQQPEASQSSEVTGSGAQA 218
Db   1544 PLAGVPGAAVADAAVGRRVGAGPGLPERGEQRPVGRRGPVGHEREVVVGGAALPARGPGG 1603

Qy    219 CESNLQGGNGGGG---------------------------------VAFYNLGMNMNNF 244
Db   1604 LRGRGRGGRGGGGGGGGGGRGPRGRGGRRRRRWRPGAGEWGAGPDSFVFFSLG------ 1657

Qy    245 QFSGGDIYGMSGGSGGGGGGATR 267
Db   1658 ---GGRGRGGRGGRGRGGGRAPR 1677
```

RESULT 12
US-08-195-152-2
; Sequence 2, Application US/08195152
; Patent No. 5679541
; GENERAL INFORMATION:
; APPLICANT: Bonini, Nancy M.
; APPLICANT: Leiserson, William M.
; APPLICANT: Benzer, Seymour
; TITLE OF INVENTION: PROGRAMMED CELL DEATH ANTAGONIST
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/195,152
; FILING DATE: 14-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-59551/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 760 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-195-152-2

Query Match 7.4%; Score 103; DB 1; Length 760;
Best Local Similarity 23.6%; Pred. No. 0.76;
Matches 64; Conservative 28; Mismatches 107; Indels 72; Gaps 11;

```
Qy     19 GGATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDS----------PNVLRSHVLE 67
Db     91 GGGGSGGGVGGGVGQCSPLGLP--PQSQPLQPTIGSLASLSGHYSNGNANPNVNSSSCSL 148

Qy     68 VTSGSDISEAVSTYATRRGCGVCI--ISGTGAVTNVTIRQPAAPAGGGVITLHGRFDILS 125
Db    149 ATASSFAQSAGSSFSTYQQAGGTSGGVSGEDGV----------VGGATVMSHWTHDGTG 197

Qy    126 LTGTALPPPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPV--------VLMAASFAN 177
Db    198 SSAAVKSESRSPG----QVHASLDNGSVAGSNLYGCSSASNPLDGGAVAVNSSAVAAAAA 253

Qy    178 AVYD------RLPIEEEETPPP----------RTTGVQQQQPE----ASQSSEVTGSGAQ 217
Db    254 AVYDGKHDYYYYNSMQQYTPPPFYSGYGTPYAAATAARQAKMEPGAAAAAAAYLTPSYAA 313

Qy    218 ACESNLQ------------GGNGGGGVAFYN 236
Db    314 SGNNNSQLYSSPYAGYNNFGQQDYGG--YYN 342
```

RESULT 13
US-09-219-849-48
; Sequence 48, Application US/09219849
; Patent No. 6150081
; GENERAL INFORMATION:
; APPLICANT: VAN HEERDE, GEORGE V.
; APPLICANT: VAN RIJN, ALEXIS C.
; APPLICANT: BOUWSTRA, JAN B.
; APPLICANT: DE WOLF, FREDERIK A.
; APPLICANT: MOOBROEK, ANDREAS
; APPLICANT: WERTEN, MARC W.T.
; APPLICANT: WIND, RICHELE D.
; APPLICANT: VAN DEN BOSCH, TANJA J.
; TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
; TITLE OF INVENTION: PREPARATION THEREOF
; FILE REFERENCE: 2728-2
; CURRENT APPLICATION NUMBER: US/09/219,849
; CURRENT FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 48
; LENGTH: 595
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: amino acid sequence
US-09-219-849-48

Query Match 7.3%; Score 101.5; DB 4; Length 595;
Best Local Similarity 23.4%; Pred. No. 0.74;
Matches 71; Conservative 19; Mismatches 118; Indels 95; Gaps 15;

```
Qy      6 SEADEAKAETTPTG--GATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDSPNVLRS 63
Db     92 SGLDGAKGDAGPAGPKGEPGSPGENGAPGQMGPRGLP-GERGRPGPP------------ 137

Qy     64 HVLEVTSGSDISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAG-GGVITLHGRFD 122
Db    138 ---------------GTAGARGNDGAVGAAGPPGPTG-----PTGPPGFPGAVGAKGEAG 177

Qy    123 ILSLTGTALP-----PPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASG-PVVLMAASFA 176
Db    178 PQGARGSEGPQGVRGEPGPPGPAG----AAGPAGNPGADGQPGAKGANGAPGIAGAPGFP 233

Qy    177 NAVYDRLPIEEEETPP----------------------PRTTGVQQQQPEASQSSEVTG 213
Db    234 GARGPSGP-QGPSGPPGPKGNSGEPGAPGNKGDTGAKGEPGATGVQGPPGPAGEEGK-RG 291

Qy    214 SGAQACESNL------QGGNGGGGVAFYNLGMNMNNFQFSGGDIYGMSGGSG-GGGGGAT 266
Db    292 ARGEPGPSGLPGPPGERGGPGSRG-------------FPGAD--GVAGPKGPSGERGAP 335

Qy    267 RPA 269
Db    336 GPA 338
```

US-09-010-928B-4

Query Match 7.7%; Score 107; DB 2; Length 907;
Best Local Similarity 26.2%; Pred. No. 0.44;
Matches 66; Conservative 13; Mismatches 85; Indels 88; Gaps 12;

```
Qy      19 GGATSSATASGS-----SSGRRPRGRPAGSKNKPKPPTIITRDSPNVLRSHVLEVT-SGS 72
           |||  |  | ||     | |    |   ||       | || |         |::|  |:
Db     627 GGAGGSGGAGGSGGAGGSGGAGGSGGVGGSGG-----TTITED---------LDITIDGA 672

Qy      73 DISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGRFDILSLTGTALP 132
           |    :|   |  | |    || |          | | | |
Db     673 DGPITISEELTISGAGG---SGPGG---------AGPGGVG------------------- 701

Qy     133 PPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVYDRLPIEEEETPP 192
           |   | ||:   ::|  |   ||:  | : : |
Db     702 -PGGSGPGGVGPGVSGPGGVGPGGSGPGGVGSGG-----------------------SG 736

Qy     193 PRTTGVQQQQPEASQSSEVTGSGAQACESNLQGGNGGGGVAFYNLGMNMNNFQFSGGDIY 252
           |   |      |  | |  | | |         ||:||    ||  | :   :  ||   |
Db     737 PGGVGPGGYGPGGSGSGGV-GPGGYG-----PGGSGG----FYGPGGSEGPYGPSG--TY 784

Qy     253 GMSGGSGGGGGG 264
           |  || | || |
Db     785 GSGGGYGPGGAG 796
```

RESULT 10
US-09-056-556-184
; Sequence 184, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND TREATM
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 184:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 460 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-09-056-556-184

Query Match 7.6%; Score 106; DB 4; Length 460;
Best Local Similarity 22.9%; Pred. No. 0.22;
Matches 55; Conservative 15; Mismatches 54; Indels 116; Gaps 11;

```
Qy      45 NKPKPPTIITRDSP------NVLRSHVLEVTSGSDISEAVSTYATRRGCGVCIISGTGAV 98
           | ||||  |  | |       ::   ::  : || ::                  |||
Db     274 NPPKPPPAIKIDPPPPPQEQGLIPGFLMPPSDGSGVT----------------PGTG-- 314

Qy      99 TNVTIRQPAAPAGGGVITLHGRFDILSLTGTALPPPAPPGAGGLTVYLAGGQGQVV-GGN 157
           ||||                          :||   || |||    |    |:   |
Db     315 ------MPAAP-------------------MVPPTGSPG-GGLP---ADTAAQLTSAGR 344

Qy     158 VAGSLIASGPVVLMAASFANAVYDRLPIEEEETPPPRTTGVQQQQPEASQSSEVTGSGAQ 217
           | :|  || | : |||
Db     345 EAAAL--SGDVAVKAASL---------------------------------------- 360

Qy     218 ACESNLQGGNGGGGVAFYNLGMNMNNFQ----FSGGDIYGM-----SGGSGGGGGGATRP 268
           || |||||    ||  :   :       ||| |:      ||:  ||||   |
Db     361 -------GGGGGGGVPSAPLGSAIGGAESVRPAGAGDIAGLGQGRAGGGAALGGGGMGMP 413
```

RESULT 11
US-07-945-283-2
; Sequence 2, Application US/07945283
; Patent No. 5352596
; GENERAL INFORMATION:
; APPLICANT: Cheung, Andrew K.
; APPLICANT: Wesley, Ronald D.
; TITLE OF INVENTION: Pseudorabies Virus Deletion Mutants
; TITLE OF INVENTION: Involving The EP0 and LLT Genes
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis P. Ribando
; STREET: 1815 No. 5352596th University Street
; CITY: Peoria
; STATE: IL
; COUNTRY: USA
; ZIP: 61604
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/945,283
; FILING DATE: 19920911
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Ribando, Curtis P
; REGISTRATION NUMBER: 27976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 309-685-4011 ext.513
; TELEFAX: 309-685-4128
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1958 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-945-283-2

Query Match 7.5%; Score 104; DB 1; Length 1958;
Best Local Similarity 20.6%; Pred. No. 2.1;
Matches 79; Conservative 36; Mismatches 142; Indels 126; Gaps 13;

```
Qy       1 MELNRSEADEAKAETTPTGGATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDS-PN 59
           : | |:   :  ||    ||     |       |   ||:|  :  :        :|:
Db    1305 LPLGRAGPGDRVAEREQRGGHLLEAGGPEGGRGAGGRGQPERAGQQALEDAAAGQDAGVR 1364

Qy      60 VLRSHVLEVTSGSDISEAVSTYATRR--GCGV----------------------CIIS 93
           |  |   :  |   ::| :     |  | ||                        |::
Db    1365 QLAGHAAGLRGGEGGADAGAEGLDGRLPGAGVRGAARVGHVGVGPAEALQDEGLLGAIVA 1424

Qy      94 ---GTGA--------------VTNVTIRQPAAPAGGGVITLHGRFD------ILSLTGT 129
              | ||              | :|  |:  |    | : : || |      : :| |
```

```
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..309
US-08-209-747-2


  Query Match             7.9%;  Score 110;  DB 1;  Length 832;
  Best Local Similarity   24.7%;  Pred. No. 0.22;
  Matches   61;  Conservative   14;  Mismatches  124;  Indels   48;  Gaps    5;

Qy      19 GGATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDSPNVLRSHVLEVTSGSDISEAV 78
           |    :| |:|: :|    || ||:                          :|:    |
Db     503 GAGAGAAAAAGAGAGGAGYGRGAGAG--------------------AGAAAGAGAGAAA 541

Qy      79 STYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGRFDILSLTGTALPPPAPPG 138
             |   | |     | ||         |   |    :  ||       |      | |
Db     542 GAGAGAGGYGGQGGYGAGARAGAAAAAGAGAGGAAGYSRGGRAGAAG-AGAGAAAGAGAG 600

Qy     139 AGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVYDRLPIEEEETPPPRTTGV 198
           |||      ||||    |  ||:  |:|     | |     | |
Db     601 AGGY-----GGQGGYGAGAGAGAAAAAG-----AGSGGAGGYGR--------------- 634

Qy     199 QQQQPEASQSSEVTGSGAQACESNLQGGNGGGGVAFYNLGMNMNNFQFSGGDIYGMSGGS 258
               |: :    |:|| |     ||| | |  |    |       :  |   |  ||
Db     635 GAGAGAAAGAGAAAGAGAGAGGYGGQGGYGAGAGAAAAAGAGAGRGGYGRGAGAGGYGGQ 694

Qy     259 GGGGGGA 265
           || | ||
Db     695 GGYGAGA 701


RESULT   8
US-08-458-298-2
; Sequence 2, Application US/08458298
; Patent No. 5756677
;  GENERAL INFORMATION:
;    APPLICANT:  Lewis, Randolph V.
;    APPLICANT:  Colgin, Mark
;    TITLE OF INVENTION:  cDNAs Encoding Minor Ampullate Spider
;    TITLE OF INVENTION:  Silk Proteins
;    NUMBER OF SEQUENCES:  56
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Birch, Stewart, Kolasch & Birch
;      STREET:  P.O. Box 747
;      CITY:  Falls Church
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22040-3487
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,298
;      FILING DATE:  02-JUN-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/209,747
;      FILING DATE:  14-MAR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murphy Jr., Gerald M.
;      REGISTRATION NUMBER:  28,977
;      REFERENCE/DOCKET NUMBER:  1447-104P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-205-8000
;      TELEFAX:  703-205-8050
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  832 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    FRAGMENT TYPE:  internal
;    ORIGINAL SOURCE:
;      ORGANISM:  N. clavipes
;      TISSUE TYPE:  minor ampullate gland
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..309
US-08-458-298-2


  Query Match             7.9%;  Score 110;  DB 1;  Length 832;
  Best Local Similarity   24.7%;  Pred. No. 0.22;
  Matches   61;  Conservative   14;  Mismatches  124;  Indels   48;  Gaps    5;

Qy      19 GGATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDSPNVLRSHVLEVTSGSDISEAV 78
           |    :| |:|: :|    || ||:                          :|:    |
Db     503 GAGAGAAAAAGAGAGGAGYGRGAGAG--------------------AGAAAGAGAGAAA 541

Qy      79 STYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGRFDILSLTGTALPPPAPPG 138
             |   | |     | ||         |   |    :  ||       |      | |
Db     542 GAGAGAGGYGGQGGYGAGARAGAAAAAGAGAGGAAGYSRGGRAGAAG-AGAGAAAGAGAG 600

Qy     139 AGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVYDRLPIEEEETPPPRTTGV 198
           |||      ||||    |  ||:  |:|     | |     | |
Db     601 AGGY-----GGQGGYGAGAGAGAAAAAG-----AGSGGAGGYGR--------------- 634

Qy     199 QQQQPEASQSSEVTGSGAQACESNLQGGNGGGGVAFYNLGMNMNNFQFSGGDIYGMSGGS 258
               |: :    |:|| |     ||| | |  |    |       :  |   |  ||
Db     635 GAGAGAAAGAGAAAGAGAGAGGYGGQGGYGAGAGAAAAAGAGAGRGGYGRGAGAGGYGGQ 694

Qy     259 GGGGGGA 265
           || | ||
Db     695 GGYGAGA 701


RESULT   9
US-09-010-928B-4
; Sequence 4, Application US/09010928B
; Patent No. 5994099
;  GENERAL INFORMATION:
;    APPLICANT:  Lewis, Randolph V
;    APPLICANT:  Hayashi, Cheryl Y
;    TITLE OF INVENTION:  EXTREMELY ELASTIC SPIDER SILK PROTEIN AND DNA
;    TITLE OF INVENTION:  CODING THEREFOR
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BIRCH, STEWART, KOLASCH & BIRCH
;      STREET:  8110 GATEHOUSE RD. SUITE 500E
;      CITY:  FALLS CHURCH
;      STATE:  VIRGINIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  22042
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/010,928B
;      FILING DATE:  22-JAN-1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murphy Jr., Gerald M
;      REGISTRATION NUMBER:  28977
;      REFERENCE/DOCKET NUMBER:  1447-109P
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  907 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
```

```
Db    198 GRQPPPAPPEQADGNAPGPQPPPVRIQDIKTENGTCPSPPQPLSPAAALGSGS--AAAVP 255

Qy     80 TYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGRFDILSLTG--TALPPPAPP 137
          :       : ||:    ::   :|                  ||| |  :| |||||
Db    256 KIESPDSSSSSLSSGSSPPGSLPSARP----------------LSLDGADSAPPPPAPS 298

Qy    138 G-----AGGLTV----YLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVYDRLPIEEE 188
          : | :|       |  |     ::  |:||      |||    :
Db    299 APPPHHSQGFSVDNIMTSLRGSPQSAAAELSSGLLASA-----AASSRAGI--------- 344

Qy    189 ETPPPRTTGVQQ------QQPEASQSSEVTGSGAQACESNLQGGNGGGGVAFYNL-GMNM 241
            ||   |            ||:|    ||     :   || || |    ||  |::
Db    345 --APPLALGAYSPGQSSLYSSPCSQTSSAGSSGGGGGGAGAAGGAGGAGTYHCNLQAMSL 402

Qy    242 NNFQFSGGDIYGMSGGSGG------------------------GGGG 264
              || : |  ||:||                           ||||
Db    403 YAAGERGGHLQGAPGGAGGSAVDNPLPDYSLPPVTSSSSSSLSHGGGG 450


RESULT   6
US-08-864-038A-3
; Sequence 3, Application US/08864038A
; Patent No. 6001592
;  GENERAL INFORMATION:
;    APPLICANT:  Kunio NAKASHIMA et al.
;    TITLE OF INVENTION:  NOVEL POLYPEPTIDE GENE cDNA, VECTOR
;    TITLE OF INVENTION:  CONTAINING SAID cDNA, HOST CELLS TRANSFORMED WITH SAID
;    TITLE OF INVENTION:  VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING
;    TITLE OF INVENTION:  SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBODY
;    TITLE OF INVENTION:  TO SAID POLYPEPTIDE
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  812-5 Hirano
;      STREET:  Isshinden
;      CITY:  Tsu-city
;      STATE:  Mie-prefecture
;      COUNTRY:  JAPAN
;      ZIP:  514-01
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.44 MB storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Microsoft Windows 95
;      SOFTWARE:  Word Perfect 6.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/864,038A
;      FILING DATE:  May 28, 1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 8-184459
;      FILING DATE:  15-July-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  C. Bruce Hamburg
;      REGISTRATION NUMBER:  22,389
;      REFERENCE/DOCKET NUMBER:  F-5610
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212)986-2340
;      TELEFAX:  (212)953-7733
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  738
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      ORGANISM:  Pinctada fucata
;      CELL TYPE:  mantle epithelial cell
;    FEATURE:
;      NAME/KEY:  peptide
;      LOCATION:  from 1 to 738
;      IDENTIFICATION METHOD:  E (by experiment)
US-08-864-038A-3


  Query Match             8.0%;  Score 112;  DB 3;  Length 738;
  Best Local Similarity   24.3%;  Pred. No. 0.12;
  Matches   75;  Conservative   26;  Mismatches  118;  Indels   90;  Gaps   11;

Qy      5 RSEADEAKAETTPTGG----------------------ATSSATASGSSSGRRP------ 36
          |: |  | |     ||                      | :||:||   || |
Db    321 RAAAAAAAAAAAAAGGGGGGGGGGGGGGGAGAAAAAAAAAASASASRQMSGIRDALGDIK 380

Qy     37 ---RGRPAGSKNKPKPPTIITR-----DSPNVLRSHVLEVTSGSDISEAVSTYATRRGCG 88
             |   | :|    |   : :      |  :||:     : | :  | : |  |:  | |
Db    381 DLLRSNGASAKASAKASAVASTKSQIDDLKDVLKDLAGLLKSSASASASASASASAGGGG 440

Qy     89 VCII------SGTGAVTNVTIRQPAAPAGGGVITLHGRFDILSLTGTALPPP-APPGAGG 141
                    | |    :     || ||||:   |        | ||    |  ||||
Db    441 GGGNGGGNGGGGGGGAGALAAALAAAGAGGGLGGGGG--------GGALAAALAAAGAGG 492

Qy    142 LTVYLAGGQGQVVG-GNVAGSLIASGPVVLMAASFANAVYDRLPIEEEETPPPRTTGVQQ 200
              || | : | | : |   |:      |||      |  :                :|
Db    493 ------GGFGGLGGLGGLGGGSAAAAAAAAAAASGGGGRALRRALR-----------RQ 534

Qy    201 QQPEASQSSEVTGSGAQACESNLQGGNGGGGVAFYNLGMNMNNFQFSGGDIYGMSGGSGG 260
           :   | ::    : | |      ||  |||:              ||   |: || ||
Db    535 MRGGGSAAAAAAAAAAAA------GGGWGGGM-------------GGGFGVGLGGGFGG 574

Qy    261 GGGGATRPA 269
          | || :  |
Db    575 GFGGGSSAA 583


RESULT   7
US-08-209-747-2
; Sequence 2, Application US/08209747
; Patent No. 5733771
;  GENERAL INFORMATION:
;    APPLICANT:  Lewis, Randolph V.
;    APPLICANT:  Colgin, Mark
;    TITLE OF INVENTION:  cDNAs Encoding Minor Ampullate Spider
;    TITLE OF INVENTION:  Silk Proteins
;    NUMBER OF SEQUENCES:  56
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Birch, Stewart, Kolasch & Birch
;      STREET:  P.O. Box 747
;      CITY:  Falls Church
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22040-3487
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/209,747
;      FILING DATE:  14-MAR-1994
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murphy Jr., Gerald M.
;      REGISTRATION NUMBER:  28,977
;      REFERENCE/DOCKET NUMBER:  1447-104P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-205-8000
;      TELEFAX:  703-205-8050
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  832 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    FRAGMENT TYPE:  internal
;    ORIGINAL SOURCE:
;      ORGANISM:  N. clavipes
;      TISSUE TYPE:  minor ampullate gland
```

```
           ||:      :  |  | :     ||   |       | |   :|:          |  |
Db    831 VLQPHLPRKVDTAAQTNSRLARPTRLSLG-----DPKASTLPRVREQQQQQQSSLHPPEP 885

Qy    110 AGGGVIT-----------LHGRFDILSLTGTALPP------------------------ 133
              |              | |     |      ||
Db    886 KSPGEYVNIEFGSGQPGYLAGPATSRSSPSVRCPPQLHPAPREETGSEEYMNMDLGPGRR 945

Qy    134 --------------PAPPGAGGL---TVYLAGGQGQVVG---GNVAGSLIASGPVVLMA 172
                       |||||:  :   |  :   :|  :    |    | : : ||
Db    946 ATWQESGGVELGRIGPAPPGSATVCRPTRSVPNSRGDYMTMQIGCPRQSYVDTSPVA--P 1003

Qy    173 ASFANAVYDRLPIEEEETPPPRTTG-VQQQQPEASQSSEVTGSGA---QACESNLQGG-N 227
           |:|:    |  |  |:   || ||        || |: ||  ||   ||  |:| ||
Db   1004 VSYADM---RTGIAAEKASLPRPTGAAPPPSSTASSSASVTPQGATAEQATHSSLLGGPQ 1060

Qy    228 GGGGV-AFYNLGMNMNNFQ--------------------FSGGDIYGMS---------GG 257
           | ||: ||  : :: |: |                    ||     | :         ||
Db   1061 GPGGMSAFTRVNLSPNHNQSAKVIRADTQGCRRRHSSETFSAPTRAGNTVPFGAGAAVGG 1120

Qy    258 SGGGGGGAT 266
           |||||||  :
Db   1121 SGGGGGGGS 1129
```

```
RESULT   4
US-09-083-351-2
; Sequence 2, Application US/09083351
; Patent No. 6087107
;  GENERAL INFORMATION:
;    APPLICANT:  Sheffield, Val C.
;    APPLICANT:  Alward, Wallace L.M.
;    APPLICANT:  Stone, Edwin M.
;    APPLICANT:  Nishimura, Darryl
;    APPLICANT:  Patil, Shiva
;    TITLE OF INVENTION:  THERAPEUTICS AND DIAGNOSTICS FOR
;    TITLE OF INVENTION:  CONGENITAL HEART DISEASE BASED ON A NOVEL HUMAN
;    TITLE OF INVENTION:  TRANSCRIPTION FACTOR
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FOLEY, HOAG & ELIOT LLP
;      STREET:  One Post Office Square
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109-2170
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/083,351
;      FILING DATE:  22-MAY-1998
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Arnold, Beth E.
;      REGISTRATION NUMBER:  35,430
;      REFERENCE/DOCKET NUMBER:  UIA-029.02
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-832-1000
;      TELEFAX:  617-832-7000
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  553 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-083-351-2
```

```
  Query Match             8.2%;  Score 113.5;  DB 3;  Length 553;
  Best Local Similarity   23.3%;  Pred. No. 0.064;
  Matches  67;  Conservative   26;  Mismatches  104;  Indels   91;  Gaps   12;
```

```
Qy     33 GRRPRGRPA----GSKNKPKPPTIITRD---------SPNVLRSHVLEVTSGSDISEAVS 79
          ||:|   |     |:   |:|| :  :|         ||    |    : |||  : ||
Db    198 GRQPPPAPPEQADGNAPGPQPPPVRIQDIKTENGTCPSPPQPLSPAAALGSGS--AAAVP 255

Qy     80 TYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGRFDILSLTG--TALPPPAPP 137
          :       : ||:    ::   :|                  ||| |  :| |||||
Db    256 KIESPDSSSSSLSSGSSPPGSLPSARP----------------LSLDGADSAPPPPAPS 298

Qy    138 G-----AGGLTV----YLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVYDRLPIEEE 188
          : | :|        |  |     ::  |:||      |||    :
Db    299 APPPHHSQGFSVDNIMTSLRGSPQSAAAELSSGLLASA-----AASSRAGI--------- 344

Qy    189 ETPPPRTTGVQQ------QQPEASQSSEVTGSGAQACESNLQGGNGGGGVAFYNL-GMNM 241
          ||   |              ||:|    ||      :   || || |    ||  |:
Db    345 --APPLALGAYSPGQSSLYSSPCSQTSSAGSSGGGGGGAGAAGGAGGAGTYHCNLQAMSL 402

Qy    242 NNFQFSGGDIYGMSGGSGG-----------------------GGGG 264
          || : |  ||:||                             ||||
Db    403 YAAGERGGHLQGAPGGAGGSAVDNPLPDYSLPPVTSSSSSSLSHGGGG 450
```

```
RESULT   5
US-09-083-352-2
; Sequence 2, Application US/09083352
; Patent No. 6207450
;  GENERAL INFORMATION:
;    APPLICANT:  Sheffield, Val C.
;    APPLICANT:  Alward, Wallace L.M.
;    APPLICANT:  Stone, Edwin M.
;    APPLICANT:  Nishimura, Darryl
;    APPLICANT:  Patil, Shiva
;    TITLE OF INVENTION:  GLAUCOMA THERAPEUTICS AND DIAGNOSTICS
;    TITLE OF INVENTION:  BASED ON A NOVEL HUMAN TRANSCRIPTION FACTOR
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FOLEY, HOAG & ELIOT LLP
;      STREET:  One Post Office Square
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109-2170
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/083,352
;      FILING DATE:  22-MAY-1998
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Arnold, Beth E.
;      REGISTRATION NUMBER:  35,430
;      REFERENCE/DOCKET NUMBER:  UIA-029.01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-832-1000
;      TELEFAX:  617-832-7000
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  553 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-083-352-2
```

```
  Query Match             8.2%;  Score 113.5;  DB 4;  Length 553;
  Best Local Similarity   23.3%;  Pred. No. 0.064;
  Matches  67;  Conservative   26;  Mismatches  104;  Indels   91;  Gaps   12;

Qy     33 GRRPRGRPA----GSKNKPKPPTIITRD---------SPNVLRSHVLEVTSGSDISEAVS 79
          ||:|   |     |:   |:|| :  :|         ||    |    : |||  : ||
```

```
Qy      59 NVLRSHV-LEVTSGSDISEAVSTYATRRG-CGVCIISGTGAVTNVTIRQPAAPAGGGVIT 116
           |   | : :  | :  | :|     || |    : |       | `  | |   | ::
Db     231 KELCKAVSVSMGLGVEALEHLSPGEQLRGDCMYAPLLGVPPAVRPTPCAPLAECKGSLL- 289

Qy     117 LHGRFDILSLTGTALPPPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVL------ 170
             |  :   |       |  ||   |  | :|: :| : : :  :|| : |
Db     290 -----DDSAGKSTEDTAEYSPFKGG---YTKGLEGESLGCSGSAAAGSSGTLELPSTLSL 341

Qy     171 -------MAASFANAVYDRLPIEEEETPPPRT-----TGVQQQQPEASQSSEVTGSGAQA 218
                 ||:: :  |   |:     |||        :: : |     |    : ||
Db     342 YKSGALDEAAAYQSRDYYNFPLALAGPPPPPPPPHPHARIKLENP-LDYGSAWAAAAAQC 400

Qy     219 CESNLQGGNGGGGV------------AFYNLGMNMNNFQFSGGDIYGMSGGSGGGGGG 264
             :|   :| |               ::: |      |    | :||  || ||||||
Db     401 RYGDLASLHGAGAAGPGSGSPSAAASSSWHTL------FTAEEGQLYGPCGGGGGGGG 453


RESULT   2
US-08-317-310A-15
; Sequence 15, Application US/08317310A
; Patent No. 5858701
;  GENERAL INFORMATION:
;    APPLICANT:  WHITE, Morris F.
;    APPLICANT:  SUN, Xiao Jian
;    APPLICANT:  PIERCE, Jacalyn H.
;    TITLE OF INVENTION:  THE IRS FAMILY OF GENES
;    NUMBER OF SEQUENCES:  64
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  28 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  ASCII text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/317,310A
;      FILING DATE:  03-OCT-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Louis Myers
;      REGISTRATION NUMBER:  35,965
;      REFERENCE/DOCKET NUMBER:  JDP-022
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1234 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FRAGMENT TYPE:  internal
US-08-317-310A-15


  Query Match             8.4%;  Score 116.5;  DB 2;  Length 1234;
  Best Local Similarity   22.8%;  Pred. No. 0.1;
  Matches   84;  Conservative   29;  Mismatches  125;  Indels  131;  Gaps   16;

Qy       5 RSEADEAKAETTPTGGATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDSPNVLRSH 64
           |  :   :   | |   :||:|:| |  |     ||  |   |                |
Db     785 RLSSSSGRLRYTATAEDSSSSTSSDSLGGGYCGARPESSLTHP-------------HHH 830

Qy      65 VLEVTSGSDISEAVSTYA-----TRRGCGVCIISGTGAVTNVTIRQ----------PAAP 109
           ||:      :  | | :     ||   |        | |   :|:          | |
Db     831 VLQPHLPRKVDTAAQTNSRLARPTRLSLG-----DPKASTLPRVREQQQQQQSSLHPPEP 885

Qy     110 AGGGVIT-----------LHGRFDILSLTGTALPP------------------------ 133
```

```
                |              | |     |     ||
Db     886 KSPGEYVNIEFGSGQPGYLAGPATSRSSPSVRCPPQLHPAPREETGSEEYMNMDLGPGRR 945

Qy     134 --------------PAPPGAGGL---TVYLAGGQGQVVG---GNVAGSLIASGPVVLMA 172
                        |||||:  :   |  :   :|  :    |    | : : ||
Db     946 ATWQESGGVELGRIGPAPPGSATVCRPTRSVPNSRGDYMTMQIGCPRQSYVDTSPVA--P 1003

Qy     173 ASFANAVYDRLPIEEEETPPPRTTG-VQQQQPEASQSSEVTGSGA---QACESNLQGG-N 227
           |:|:    |  |  |:   || ||        || |: ||  ||   ||  |:| ||
Db    1004 VSYADM---RTGIAAEKASLPRPTGAAPPPSSTASSSASVTPQGATAEQATHSSLLGGPQ 1060

Qy     228 GGGGV-AFYNLGMNMNNFQ-------------------FSGGDIYGMS---------GG 257
           | ||: ||  : :: |: |                  ||     | :           ||
Db    1061 GPGGMSAFTRVNLSPNHNQSAKVIRADTQGCRRRHSSETFSAPTRAGNTVPFGAGAAVGG 1120

Qy     258 SGGGGGGAT 266
           |||||||  :
Db    1121 SGGGGGGGS 1129


RESULT   3
PCT-US95-13041-15
; Sequence 15, Application PC/TUS9513041
;  GENERAL INFORMATION:
;    APPLICANT:  WHITE, Morris F.
;    APPLICANT:  SUN, Xiao Jian
;    APPLICANT:  PIERCE, Jacalyn H.
;    TITLE OF INVENTION:  THE IRS FAMILY OF GENES
;    NUMBER OF SEQUENCES:  63
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  60 State Street, Suite 510
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  ASCII text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/13041
;      FILING DATE:  Herewith
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/317,310
;      FILING DATE:  03-OCT-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Louis Myers
;      REGISTRATION NUMBER:  35,965
;      REFERENCE/DOCKET NUMBER:  JDP-022PC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1234 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FRAGMENT TYPE:  internal
PCT-US95-13041-15


  Query Match             8.4%;  Score 116.5;  DB 5;  Length 1234;
  Best Local Similarity   22.8%;  Pred. No. 0.1;
  Matches   84;  Conservative   29;  Mismatches  125;  Indels  131;  Gaps   16;

Qy       5 RSEADEAKAETTPTGGATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDSPNVLRSH 64
           |  :   :   | |   :||:|:| |  |     ||  |   |                |
Db     785 RLSSSSGRLRYTATAEDSSSSTSSDSLGGGYCGARPESSLTHP-------------HHH 830

Qy      65 VLEVTSGSDISEAVSTYA-----TRRGCGVCIISGTGAVTNVTIRQ----------PAAP 109
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2002, 21:34:28 ; Search time 46.14 Seconds
(without alignments)
142.933 Million cell updates/sec

Title: US-09-823-676-2
Perfect score: 1392
Sequence: 1 MELNRSEADEAKAETTPTGG.....IYGMSGGSGGGGGGATRPAF 270

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	118.5	8.5	918	4	US-09-041-886-11	Sequence 11, Appl
2	116.5	8.4	1234	2	US-08-317-310A-15	Sequence 15, Appl
3	116.5	8.4	1234	5	PCT-US95-13041-15	Sequence 15, Appl
4	113.5	8.2	553	3	US-09-083-351-2	Sequence 2, Appli
5	113.5	8.2	553	4	US-09-083-352-2	Sequence 2, Appli
6	112	8.0	738	3	US-08-864-038A-3	Sequence 3, Appli
7	110	7.9	832	1	US-08-209-747-2	Sequence 2, Appli
8	110	7.9	832	1	US-08-458-298-2	Sequence 2, Appli
9	107	7.7	907	2	US-09-010-928B-4	Sequence 4, Appli
10	106	7.6	460	4	US-09-056-556-184	Sequence 184, App
11	104	7.5	1958	1	US-07-945-283-2	Sequence 2, Appli
12	103	7.4	760	1	US-08-195-152-2	Sequence 2, Appli
13	101.5	7.3	595	4	US-09-219-849-48	Sequence 48, Appl
14	101.5	7.3	595	4	US-09-219-849-50	Sequence 50, Appl
15	101.5	7.3	822	4	US-09-219-849-49	Sequence 49, Appl
16	101	7.3	1185	4	US-09-041-886-23	Sequence 23, Appl
17	100.5	7.2	412	4	US-09-027-064-4	Sequence 4, Appli
18	100.5	7.2	412	4	US-09-271-815-4	Sequence 4, Appli
19	99.5	7.1	378	2	US-08-986-217-6	Sequence 6, Appli
20	99	7.1	485	2	US-08-749-391-2	Sequence 2, Appli
21	99	7.1	485	4	US-09-390-200-2	Sequence 2, Appli
22	98	7.0	318	4	US-09-060-756-727	Sequence 727, App
23	98	7.0	334	4	US-09-060-756-728	Sequence 728, App
24	98	7.0	2055	1	US-08-175-155-46	Sequence 46, Appl
25	98	7.0	2055	1	US-08-477-509B-81	Sequence 81, Appl
26	98	7.0	2055	2	US-08-707-237A-52	Sequence 52, Appl
27	98	7.0	2055	3	US-08-482-085B-81	Sequence 81, Appl
28	98	7.0	2055	4	US-09-444-791A-81	Sequence 81, Appl
29	97.5	7.0	440	3	US-09-100-664A-2	Sequence 2, Appli
30	97.5	7.0	440	3	US-09-100-664A-3	Sequence 3, Appli
31	97.5	7.0	440	3	US-09-100-664A-4	Sequence 4, Appli
32	97.5	7.0	1011	1	US-08-477-509B-94	Sequence 94, Appl
33	97.5	7.0	1011	3	US-08-482-085B-94	Sequence 94, Appl
34	97.5	7.0	1011	4	US-09-444-791A-94	Sequence 94, Appl
35	97.5	7.0	1026	1	US-08-194-290-7	Sequence 7, Appli
36	97.5	7.0	1170	1	US-08-175-155-59	Sequence 59, Appl
37	97.5	7.0	1170	2	US-08-707-237A-66	Sequence 66, Appl
38	97	7.0	400	4	US-09-086-010-2	Sequence 2, Appli
39	96.5	6.9	380	2	US-09-026-587-4	Sequence 4, Appli
40	96.5	6.9	380	2	US-09-227-420-4	Sequence 4, Appli
41	96	6.9	652	1	US-08-050-684-2	Sequence 2, Appli
42	96	6.9	652	1	US-08-582-719-2	Sequence 2, Appli
43	96	6.9	1038	1	US-07-609-716-36	Sequence 36, Appl
44	96	6.9	1038	4	US-08-475-411A-36	Sequence 36, Appl
45	96	6.9	1038	4	US-08-478-029A-36	Sequence 36, Appl

ALIGNMENTS

```
RESULT 1
US-09-041-886-11
; Sequence 11, Application US/09041886
; Patent No. 6235872
;  GENERAL INFORMATION:
;    APPLICANT:  Bredesen, Dale E.
;    APPLICANT:  Rabizadeh, Sharroz
;    TITLE OF INVENTION:  Proapoptotic Peptides, Dependence
;    TITLE OF INVENTION:  Polypeptides and Methods of Use
;    NUMBER OF SEQUENCES:  72
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell & Flores LLP
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/041,886
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LJ 2626
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  918 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-041-886-11


  Query Match             8.5%;  Score 118.5;  DB 4;  Length 918;
  Best Local Similarity   23.1%;  Pred. No. 0.046;
  Matches   69;  Conservative   36;  Mismatches  129;  Indels   65;  Gaps   11;

Qy       1 MELNRSEADEAKAETTPTGGATSSATASGSSSG--RRPRGRPAGSKNKPKPPTIITRDSP 58
           |:| : :  || :|               |||||  |   | |  ||:     |    |:
Db     185 MQLLQQQQQEAVSE-------------GSSSGRAREASGAPTSSKDNYLGGTSTISDNA 230
```

```
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             40.9%;  Score 570;  DB 21;  Length 239;
  Best Local Similarity   46.9%;  Pred. No. 2.2e-38;
  Matches  119;  Conservative   40;  Mismatches   49;  Indels   46;  Gaps    5;

Qy       3 LNRSEADEAKAETTPTGGATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDSPNVLR 62
           :|:|: ::   |  |  ||            |||||| |||||| |  :|||||| ||
Db       1 mnqsqdndqdeeddpregavevv-------nrrprgrppgsknkpkapifvtrdspnalr 53

Qy      63 SHVLEVTSGSDISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGRFD 122
           |||||:: |||::: :: :: ||  |||::||||:| |||:|| |||  |||::| |||:
Db      54 shvleisdgsdvadtiahfsrrrqrgvcvlsgtgsvanvtlrqaaap--ggvvslqgrfe 111

Qy     123 ILSLTGTALPPPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVYDR 182
           ||||||  || |:|||: |||||||| |||||||:| | |:| | |:::||:|:|| |:|
Db     112 ilsltgaflpgpsppgstgltvylagvqggqvvggsvvgpllaigsvmviaatfsnatyer 171

Qy     183 LPIEEEETPPPRTTGVQQQQPEASQSSEVTGSGAQACESNLQGGNGGGGVAFYNLGMNMN 242
           ||:|||                    | |::           ||| :   :| |:
Db     172 lpmeeee---------------------dgggsrqih--------gggdspprigsnl- 200

Qy     243 NFQFSGGDIYGMSG 256
                 |: ||:|
Db     201 ------pdlsgmag 208


Search completed: July 10, 2002, 22:20:18
Job time: 3611 sec
```

XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.

PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.

PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 41.1%; Score 572; DB 21; Length 316;
Best Local Similarity 47.8%; Pred. No. 2.2e-38;
Matches 120; Conservative 44; Mismatches 69; Indels 18; Gaps 6;

```
Qy       5 RSEADEAKAETTPTGGATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDSPNVLRSH 64
           :|: ::   :  |   |    : || |:||||||| |||||| | ::|::||| |:||
Db      69 qsqeeqnsrdeqp---avepgsgsg-stgrrprgrppgsknkpkspvvvtkespnslqsh 124

Qy      65 VLEVTSGSDISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGRFDIL 124
           |||: :|:|:::|::: :| ||| || ::||:| |||||:|||  | |||::| |:|:||
Db     125 vleiatgadvaeslnafarrrgrgvsvlsgsglvtnvtlrqpa--asggvvslrgqfeil 182

Qy     125 SLTGTALPPP-APPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVYDRL 183
           |: |  ||   :|  | |||:|||| ||||||| ||| |||||||:::||:| || |:||
Db     183 smcgaflptsgspaaaagltiylagaqggvvgggvagpliasgpviviaatfcnatyerl 242

Qy     184 PIEEEETPPPRTTGVQQQQPEASQSSEVTGSGAQACESNLQGGNGGGGVAFYNLGMNM-- 241
           |||||:         ||:||   :  :          ||   |  |      ||:  |
Db     243 pieeeq---------qqeqplqledgkkqkeenddnesgnngnegsmqppmynmppnfip 293

Qy     242 NNFQFSGGDIY 252
           |  | :  |:|
Db     294 nghqmaqhdvy 304
```

RESULT 15
AAG41483
ID AAG41483 standard; Protein; 239 AA.
XX
AC AAG41483;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 51617.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 41.1%; Score 572; DB 21; Length 315;
Best Local Similarity 47.8%; Pred. No. 2.2e-38;
Matches 120; Conservative 44; Mismatches 69; Indels 18; Gaps 6;

```
Qy       5 RSEADEAKAETTPTGGATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDSPNVLRSH 64
           :|: ::   :  |   |    : || |:||||||| |||||||  | ::|::||| |:||
Db      68 qsqeeqnsrdeqp---avepgsgsg-stgrrprgrppgsknkpkspvvvtkespnslqsh 123

Qy      65 VLEVTSGSDISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGRFDIL 124
           |||: :|:|::|::: :| ||| || ::||:| |||||:|||| | |||::| |:|:||
Db     124 vleiatgadvaeslnafarrrgrgvsvlsgsglvtnvtlrqpa--asggvvslrgqfeil 181

Qy     125 SLTGTALPPP-APPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVYDRL 183
           |: |  ||   :|  | |||:|||| ||||||| ||| |||||||:::||:| || |:||
Db     182 smcgaflptsgspaaaagltiylagaqggvvgggvagpliasgpviviaatfcnatyerl 241

Qy     184 PIEEEETPPPRTTGVQQQQPEASQSSEVTGSGAQACESNLQGGNGGGGVAFYNLGMNM-- 241
           |||||:         ||:||   :  :         ||   |  |      ||:  |
Db     242 pieeeq---------qqeqplqledgkkqkeenddnesgnngnegsmqppmynmppnfip 292

Qy     242 NNFQFSGGDIY 252
           |  | :  |:|
Db     293 nghqmaqhdvy 303
```

RESULT 14
AAG21495
ID AAG21495 standard; Protein; 316 AA.
XX
AC AAG21495;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 24071.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX

PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.

PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.

PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 41.1%; Score 572; DB 21; Length 310;
Best Local Similarity 47.8%; Pred. No. 2.1e-38;
Matches 120; Conservative 44; Mismatches 69; Indels 18; Gaps 6;

```
Qy       5 RSEADEAKAETTPTGGATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDSPNVLRSH 64
           :|: ::   :  |   |    : || |:||||||| |||||| | ::|::||| |:||
Db      63 qsqeeqnsrdeqp---avepgsgsg-stgrrprgrppgsknkpkspvvvtkespnslqsh 118

Qy      65 VLEVTSGSDISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGRFDIL 124
           |||: :|:|::|::: :| ||| || ::||:| |||||:|||| | |||::| |:|:||
Db     119 vleiatgadvaeslnafarrrgrgvsvlsgsglvtnvtlrqpa--asggvvslrgqfeil 176

Qy     125 SLTGTALPPP-APPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVYDRL 183
           |: |  ||   :|  | |||:|||| ||||||| ||| |||||||:::||:| || |:||
Db     177 smcgaflptsgspaaaagltiylagaggqvvgggvagpliasgpviviaatfcnatyerl 236

Qy     184 PIEEEETPPPRTTGVQQQQPEASQSSEVTGSGAQACESNLQGGNGGGGVAFYNLGMNM-- 241
           |||||:         ||:||   :  :         ||   |  |      ||:  |
Db     237 pieeeq---------qqeqplqledgkkqkeenddnesgnngnegsmqppmynmppnfip 287

Qy     242 NNFQFSGGDIY 252
           |  | :  |:|
Db     288 nghqmaqhdvy 298
```

RESULT 13
AAG54090
ID AAG54090 standard; Protein; 315 AA.
XX
AC AAG54090;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 68932.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.

```
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             41.1%;  Score 572;  DB 21;  Length 310;
  Best Local Similarity   47.8%;  Pred. No. 2.1e-38;
  Matches  120;  Conservative   44;  Mismatches   69;  Indels   18;  Gaps    6;

Qy       5 RSEADEAKAETTPTGGATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDSPNVLRSH 64
           :|: ::   :  |   |    : || |:||||||| |||||| | ::|::||| |:||
Db      63 qsqeeqnsrdeqp---avepgsgsg-stgrrprgrppgsknkpkspvvvtkespnslqsh 118

Qy      65 VLEVTSGSDISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGRFDIL 124
           |||: :|:|::|::: :| ||| || ::||:| |||||:|||| | |||::| |:|:||
Db     119 vleiatgadvaeslnafarrrgrgvsvlsgsglvtnvtlrqpa--asggvvslrgqfeil 176

Qy     125 SLTGTALPPP-APPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVYDRL 183
           |: |  ||   :|  | |||:|||| ||||||| ||| |||||||:::||:| || |:||
Db     177 smcgaflptsgspaaaagltiylagaqgqvvgggvagpliasgpviviaatfcnatyerl 236

Qy     184 PIEEEETPPPRTTGVQQQQPEASQSSEVTGSGAQACESNLQGGNGGGGVAFYNLGMNM-- 241
           |||||:         ||:||   :  :          ||   |  |      ||:  |
Db     237 pieeeq---------qqeqplqledgkkqkeenddnesgnngnegsmqppmynmppnfip 287

Qy     242 NNFQFSGGDIY 252
           |  | :  |:|
Db     288 nghqmaqhdvy 298


RESULT  12
AAG54091
ID   AAG54091 standard; Protein; 310 AA.
XX
AC   AAG54091;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 68933.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   20-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0144814.
PR   21-JUL-1999;   99US-0145086.
PR   21-JUL-1999;   99US-0145088.
PR   22-JUL-1999;   99US-0145085.
```

PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.

PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 41.1%; Score 572; DB 21; Length 272;
Best Local Similarity 47.8%; Pred. No. 1.8e-38;
Matches 120; Conservative 44; Mismatches 69; Indels 18; Gaps 6;

```
Qy       5 RSEADEAKAETTPTGGATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDSPNVLRSH 64
           :|: ::   :  |   |    : || |:||||||| |||||| | ::|::||| |:||
Db      25 qsqeeqnsrdeqp---avepgsgsg-stgrrprgrppgsknkpkspvvvtkespnslqsh 80

Qy      65 VLEVTSGSDISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGRFDIL 124
           |||: :|:|::|::: :| ||| || ::||:| |||||:|||| | |||::| |:|:||
Db      81 vleiatgadvaeslnafarrrgrgvsvlsgsglvtnvtlrqpa--asggvvslrgqfeil 138

Qy     125 SLTGTALPPP-APPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVYDRL 183
           |: |  ||   :|  | |||:|||| ||||||| ||| |||||||:::||:| || |:||
Db     139 smcgaflptsgspaaaagltiylagaqggvvgggvagpliasgpviviaatfcnatyerl 198

Qy     184 PIEEEETPPPRTTGVQQQQPEASQSSEVTGSGAQACESNLQGGNGGGGVAFYNLGMNM-- 241
           |||||:         ||:||   :  :          ||   |  |      ||:  |
Db     199 pieeeq--------qqeqplqledgkkqkeenddnesgnngnegsmqppmynmppnfip 249

Qy     242 NNFQFSGGDIY 252
           |  | :  |:|
Db     250 nghqmaqhdvy 260
```

RESULT 11
AAG21496
ID AAG21496 standard; Protein; 310 AA.
XX
AC AAG21496;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 24072.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.

```
Db     81 vleiatgadvaeslnafarrrgrgvsvlsgsglvtnvtlrqpa--asggvvslrgqfeil 138

Qy    125 SLTGTALPPP-APPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVYDRL 183
          |: |  ||   :|  | |||:|||| ||||||| ||| |||||||:::||:| || |:||
Db    139 smcgaflptsgspaaaagltiylagaqggvvgggvagpliasgpviviaatfcnatyerl 198

Qy    184 PIEEEETPPPRTTGVQQQQPEASQSSEVTGSGAQACESNLQGGNGGGGVAFYNLGMNM-- 241
          |||||:         ||:||   :  :          ||   |  |      ||:  |
Db    199 pieeeq---------qqeqplqledgkkqkeenddnesgnngnegsmqppmynmppnfip 249

Qy    242 NNFQFSGGDIY 252
          |  | :  |:|
Db    250 nghqmaqhdvy 260


RESULT  10
AAG54092
ID   AAG54092 standard; Protein; 272 AA.
XX
AC   AAG54092;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 68934.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   20-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0144814.
PR   21-JUL-1999;   99US-0145086.
PR   21-JUL-1999;   99US-0145088.
PR   22-JUL-1999;   99US-0145085.
PR   22-JUL-1999;   99US-0145087.
PR   22-JUL-1999;   99US-0145089.
PR   22-JUL-1999;   99US-0145192.
PR   23-JUL-1999;   99US-0145145.
PR   23-JUL-1999;   99US-0145218.
PR   23-JUL-1999;   99US-0145224.
PR   26-JUL-1999;   99US-0145276.
PR   27-JUL-1999;   99US-0145913.
PR   27-JUL-1999;   99US-0145918.
PR   27-JUL-1999;   99US-0145919.
PR   28-JUL-1999;   99US-0145951.
PR   02-AUG-1999;   99US-0146386.
PR   02-AUG-1999;   99US-0146388.
PR   02-AUG-1999;   99US-0146389.
PR   03-AUG-1999;   99US-0147038.
PR   04-AUG-1999;   99US-0147204.
PR   04-AUG-1999;   99US-0147302.
PR   05-AUG-1999;   99US-0147192.
```

PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 41.1%; Score 572; DB 21; Length 272;
Best Local Similarity 47.8%; Pred. No. 1.8e-38;
Matches 120; Conservative 44; Mismatches 69; Indels 18; Gaps 6;

```
Qy       5 RSEADEAKAETTPTGGATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDSPNVLRSH 64
             :|: ::   :  |   |    : || |:||||||||| ||||||| | ::|::||| |:||
Db      25 qsqeeqnsrdeqp---avepgsgsg-stgrrprgrppgsknkpkspvvvtkespnslqsh 80

Qy      65 VLEVTSGSDISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGRFDIL 124
           |||: :|:|::|::: :| ||| || ::||:| |||||:|||| | |||::| |:|:||
```

XX
PF 14-NOV-2000; 2000WO-US31414.
XX
PR 17-NOV-1999; 99US-0166228.
PR 17-APR-2000; 2000US-0197899.
PR 22-AUG-2000; 2000US-0227439.
XX
PA (MEND-) MENDEL BIOTECHNOLOGY INC.
PA (JIAN/) JIANG C.
PA (HEAR/) HEARD J.
PA (PINE/) PINEDA O.
PA (PILG/) PILGRIM M.
PA (ADAM/) ADAM L.
PA (RIEC/) RIECHMANN J L.
PA (YUGG/) YU G.
PA (SAMA/) SAMAHA R.
XX
PI Jiang C, Heard J, Pineda O, Pilgrim M, Adam L, Riechmann JL;
PI Yu G, Samaha R;
XX
DR WPI; 2001-335977/35.
DR N-PSDB; AAD06672.
XX
PT Nucleic acids encoding plant transcription factor polypeptides, useful
PT for altering the sugar sensing characteristics of plants and increasing
PT yield, e.g. corn, potato and cotton plants -
XX
PS Claim 4; Page 144-145; 151pp; English.
XX
CC The patent relates to polynucleotides encoding 35 plant transcription
CC factors which may be used to modify phenotype associated with a plant's
CC sugar sensing characteristics and increasing yield when their expression
CC level is altered. Sugars are central regulatory molecules that control
CC aspects of physiology, metabolism and development. Therefore the cDNAs
CC and proteins of the invention are useful for modifying the growth and
CC germination rates of plants, photosynthesis, glyoxylate metabolism,
CC respiration, starch and sucrose synthesis and degradation, pathogen
CC response, wounding response, cell cycle regulation, pigmentation,
CC flowering and senescence of plants and for modifying sink-source
CC relationships in seeds, tubers, roots, and other storage organs leading
CC to an increase in yield. The transcription factor polynucleotides and
CC polypeptides may be used to alter the structure and developmental
CC characteristics of plants such as soybean, wheat, corn, potato, cotton,
CC rice, oilseed, sunflower, alfalfa, sugarcane, turf, banana, blackberry,
CC blueberry, strawberry, raspberry, cantaloupe, carrot, cauliflower,
CC coffee, cucumber, eggplant, grapes, honey dew, lettuce, mango, melon,
CC onion, papaya, peas, peppers, pineapple, spinach, squash, sweet corn,
CC tobacco, tomato, watermelon, rosaceous fruits and/or vegetable brassicas.
CC The present sequence is a homolog of Arabidopsis thaliana transcription
CC factor.
XX
SQ Sequence 339 AA;

Query Match 44.9%; Score 625.5; DB 22; Length 339;
Best Local Similarity 49.2%; Pred. No. 1.1e-42;
Matches 129; Conservative 28; Mismatches 48; Indels 57; Gaps 3;

```
Qy       4 NRSEADEAKAETTPTGGATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDSPNVLRS 63
           |:  | |  :  |  |    ||   |||||||||||||| | |||||| | ||:
Db      88 ntnsgsegkemslhggeggsggggsgeqmtrrprgrpagsknkpkapiiitrdsanalrt 147

Qy      64 HVLEVTSGSDISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGRFDI 123
           ||:|:  | || : ::|:| ||  |||::||||:||||||||| :|  | |::|||||:|
Db     148 hvmeigdgcdivdcmatfarrrqrgvcvmsgtgsvtnvtirqpgspp-gsvvslhgrfei 206

Qy     124 LSLTGTALPPPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVYDRL 183
           |||:|: |||||| | ||:|||||||||||||:| | |: |||||:|||||:|| |:||
Db     207 lslsgsflpppappaatglsvylaggqgqvvggsvvgpllcsgpvvvmaasfsnaayerl 266

Qy     184 PIEEEETPPPRTTGVQQQQPEASQSSEVTGSGAQACESNLQGGNGGGGVAFYNLGMNMNN 243
           |:||:|   |                          :||| ||||
Db     267 pleedemqtp----------------------------vqgggggggg---------- 285
```

Qy 244 FQFSGGDIYGMSGGSGGGGGGA 265
|||| |:
Db 286 --------------ggggmgs 292

RESULT 9
AAG21497
ID AAG21497 standard; Protein; 272 AA.
XX
AC AAG21497;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 24073.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.

PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 45.8%; Score 637.5; DB 21; Length 315;
Best Local Similarity 50.0%; Pred. No. 1.1e-43;
Matches 132; Conservative 37; Mismatches 58; Indels 37; Gaps 5;

```
Qy      17 PTGGATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDSPNVLRSHVLEVTSGSDISE 76
           |  ||  : |       ||||||||||||||||  :|||||| |:|||:|: ||:|: |
Db      70 pregaveapt-------rrprgrpagsknkpkppifvtrdspnalkshvmeiasgtdvie 122

Qy      77 AVSTYATRRGCGVCIISGTGAVTNVTIRQP------AAPAGGGVITLHGRFDILSLTGTA 130
              ::|:| ||  |:||:|| | | |||:|||      ||| |  |: | |||:||||||:
Db     123 tlatfarrrqrgicilsgngtvanvtlrqpstaavaaapggaavlalqgrfeilsltgsf 182

Qy     131 LPPPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVYDRLPIEEEET 190
           || ||||||: |||:|||||||||||||:| | |:|:|||:|:||:|:|| |:|||:||||
Db     183 lpgpappgstgltiylaggqgqvvggsvvgplmaagpvmliaatfsnatyerlpleeeea 242

Qy     191 PP----PRTTGVQQQQPEASQSSEVTGSGAQACESNLQGGNGGGGVAFYNLGMNMNNFQF 246
                  : ||   |          | |     |   ||:|  |:  ||:  |: :
Db     243 aergggggsggvvpgq---------lggggsplssgagggdgnqglpvynmpgnlvs--- 290

Qy     247 SGGDIYGMSGGSGGGGGGATRPAF 270
                 :|||||||  : : |:
Db     291 --------nggsggggqmsgqeay 306
```

RESULT 8
AAE02571
ID AAE02571 standard; Protein; 339 AA.
XX
AC AAE02571;
XX
DT 10-AUG-2001 (first entry)
XX
DE A. thaliana transcription factor G1068 homolog, G2657.
XX
KW Plant transcription factor; phenotype; sugar sensing characteristic;
KW transgenic plant; plant yield; growth; germination; photosynthesis;
KW glyoxylate metabolism; respiration; pathogen response; wounding response;
KW cell cycle regulation; pigmentation; flowering; senescence; physiology;
KW storage organ; metabolism.
XX
OS Arabidopsis thaliana.
XX
FH Key Location/Qualifiers
FT Domain 116..129
FT /note= "Conserved domain"
XX
PN WO200135725-A1.
XX
PD 25-MAY-2001.

PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 45.8%; Score 637.5; DB 21; Length 282;
Best Local Similarity 50.0%; Pred. No. 9.5e-44;
Matches 132; Conservative 37; Mismatches 58; Indels 37; Gaps 5;

```
Qy     17 PTGGATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDSPNVLRSHVLEVTSGSDISE 76
          |  ||  : |       ||||||||||||||||  :|||||| |:|||:|: ||:|: |
Db     37 pregaveapt-------rrprgrpagsknkpkppifvtrdspnalkshvmeiasgtdvie 89

Qy     77 AVSTYATRRGCGVCIISGTGAVTNVTIRQP------AAPAGGGVITLHGRFDILSLTGTA 130
            ::|:| ||  |:||:|| | | |||:|||      ||| |  |: | |||:|||||||:
Db     90 tlatfarrrqrgicilsgngtvanvtlrqpstaavaaapggaavlalqgrfeilsltgsf 149

Qy    131 LPPPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVYDRLPIEEEET 190
          || |||||: |||:|||||||||||||:| | |:|:|||:|:||:|:|| |:|||:||||
Db    150 lpgpappgstgltiylaggqggvvggsvvgplmaagpvmliaatfsnatyerlpleeeea 209

Qy    191 PP----PRTTGVQQQQPEASQSSEVTGSGAQACESNLQGGNGGGGVAFYNLGMNMNNFQF 246
                : ||   |          | |     |   ||:|  |:  ||:  |: :
Db    210 aergggggsggvvpgq---------lggggsplssgagggdgnqglpvynmpgnlvs--- 257

Qy    247 SGGDIYGMSGGSGGGGGGATRPAF 270
                :|||||||  : : |:
Db    258 --------nggsggggqmsgqeay 273
```

RESULT 7
AAG10648
ID AAG10648 standard; Protein; 315 AA.
XX
AC AAG10648;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 9054.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2002, 19:01:52 ; Search time 1922.93 Seconds
(without alignments)
10599.682 Million cell updates/sec

Title: US-09-823-676-1
Perfect score: 974
Sequence: 1 ccccccgacctgcctctaca.....cgaaatgcgaatattaggtt 974

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	952.8	97.8	86710	8	ATF23E12	AL022604 Arabidops
2	952.8	97.8	197859	8	ATCHRIV83	AL161587 Arabidops
3	212.4	21.8	1323	8	AF194974	AF194974 Arabidops
c 4	212.4	21.8	103192	8	AC007369	AC007369 Arabidops
5	208.6	21.4	88411	8	AC004667	AC004667 Arabidops
6	205.2	21.1	122871	8	AC002387	AC002387 Arabidops
7	201.8	20.7	100806	8	AC015450	AC015450 Arabidops
c 8	196.8	20.2	107600	8	ATF16J13	AL049638 Arabidops
c 9	196.8	20.2	190026	8	ATCHRIV33	AL161533 Arabidops
10	191.6	19.7	197976	8	ATCHRIV39	AL161539 Arabidops
11	191.6	19.7	206606	8	ATFCA1	Z97336 Arabidopsis
12	190.4	19.5	1161	8	AF446359	AF446359 Arabidops
c 13	190.4	19.5	81835	8	ATFCA9	Z97344 Arabidopsis
c 14	190.4	19.5	198067	8	ATCHRIV47	AL161547 Arabidops
15	186.4	19.1	89350	8	ATT12H17	AL021635 Arabidops
16	186.4	19.1	193550	8	ATCHRIV58	AL161558 Arabidops
c 17	183	18.8	129845	2	AP004165	AP004165 Oryza sat
18	180.2	18.5	177418	2	OSJN00182	AL662981 Oryza sat
19	179.4	18.4	107865	8	ATT4C21	AL162295 Arabidops
c 20	178.8	18.4	103240	8	ATT22E16	AL132975 Arabidops
21	177.2	18.2	95310	8	ATAC011437	AC011437 Arabidops
22	174	17.9	160363	2	AP003526	AP003526 Oryza sat
c 23	170.4	17.5	116205	8	AC006931	AC006931 Arabidops
c 24	166	17.0	156834	2	AP004757	AP004757 Oryza sat
25	161.8	16.6	134712	2	AP003891	AP003891 Oryza sat
26	161.8	16.6	135769	2	AP004587	AP004587 Oryza sat
27	160.8	16.5	108763	2	AP004020	AP004020 Oryza sat
c 28	148.8	15.3	147666	2	AP004635	AP004635 Oryza sat
29	148.8	15.3	152381	2	AP004698	AP004698 Oryza sat
c 30	127.6	13.1	12319	6	E64635	E64635 Method for
31	127.6	13.1	71807	8	AB025613	AB025613 Arabidops
32	116.8	12.0	145847	2	AP004319	AP004319 Oryza sat
33	111.6	11.5	140466	8	AP003683	AP003683 Oryza sat
c 34	108.2	11.1	111686	8	F14L17	AC012188 Sequence
35	101.2	10.4	279	8	AF184830	AF184830 Arabidops
c 36	89.2	9.2	140327	2	AP003938	AP003938 Oryza sat
c 37	87.8	9.0	169656	2	AP004680	AP004680 Oryza sat
38	69.8	7.2	167764	2	AP004303	AP004303 Oryza sat
c 39	50.8	5.2	125020	9	AF429315	AF429315 Homo sapi
40	45.2	4.6	590	8	HVU234403	AJ234403 Hordeum v
41	45.2	4.6	124050	8	AF474373	AF474373 Hordeum v
c 42	44.6	4.6	7218	6	I66494	I66494 Sequence 14
43	43.8	4.5	125020	9	AF429315	AF429315 Homo sapi
44	42.2	4.3	1137	8	AMA132349	AJ132349 Antirrhin
45	42	4.3	1694	8	ATAJ4119	AJ224119 Arabidops

ALIGNMENTS

RESULT 1
ATF23E12/c
LOCUS ATF23E12 86710 bp DNA linear PLN 01-APR-1999
DEFINITION Arabidopsis thaliana DNA chromosome 4, BAC clone F23E12 (ESSA project).
ACCESSION AL022604
VERSION AL022604.1 GI:3080406
KEYWORDS .
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 86710)
AUTHORS Bevan,M., Hilbert,H., Braun,M., Holzer,E., Brandt,A., Duesterhoeft,A., Hoheisel,J., Mewes,H.W., Mayer,K.F.X. and Schueller,C.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 86710)
AUTHORS EU Arabidopsis sequencing,project.
TITLE Direct Submission
JOURNAL Submitted (01-APR-1999) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:

```
            schuelle@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
COMMENT     Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
            and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.
FEATURES             Location/Qualifiers
     source          1. .86710
                     /organism="Arabidopsis thaliana"
                     /variety="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="4"
     gene            173. .1890
                     /gene="F23E12.10"
     exon            173. .374
                     /gene="F23E12.10"
                     /number=1
     CDS             join(<173. .374,456. .919,1089. .1610,1678. .1890)
                     /gene="F23E12.10"
                     /codon_start=1
                     /product="hypothetical protein (fragment)"
                     /protein_id="CAA18726.1"
                     /db_xref="GI:3080407"
                     /db_xref="SPTREMBL:O65486"
                     /translation="FNAAVAGCFFAVESVLWPSSSTDSSTSLPNTTSMVILSAVTASV
                     VSEIGLGSEPAFKVPDYDFRSPGELPLYLLLGALCGLVSLALSRCTSSMTSAVDSLNK
                     DAGIPKAVFPVMGGLSVGIIALVYPEVLYWGFQNVDILLEKRPFVKGLSADLLLQLVA
                     VKIAATAWCRASGLVGGYYAPSLFIGGAAGMAYGKFIGLALAQNPDFNLSILEVASPQ
                     AYGLVGMAATLAGVCQVPLTAVLLLFELTQDYRIVLPLLGAVGMSSWITSGQSKRQET
                     RETKETRKRKSQEAVQSLTSSDDESSTNNLCEVESSLCLDDSLNQSEELPKSIFVSEA
                     MRTRFATVMMSTSLEEALTRMLIEKQSCALIVDPDNIFLGILTLSDIQEFSKARKEGN
                     NRPKDIFVNDICSRSGGKCKVPWTVTPDMDLLAAQTIMNKHELSHVAVVSGSIDAPRI
                     HPVGVLDRECITLTRR"
     intron          375. .455
                     /gene="F23E12.10"
                     /number=1
     exon            456. .919
                     /gene="F23E12.10"
                     /number=2
     intron          920. .1088
                     /gene="F23E12.10"
                     /number=2
     exon            1089. .1610
                     /gene="F23E12.10"
                     /number=3
     intron          1611. .1677
                     /gene="F23E12.10"
                     /number=3
     exon            1678. .1890
                     /gene="F23E12.10"
                     /number=4
     exon            3124. .3241
                     /gene="F23E12.20"
                     /number=1
     gene            3124. .4428
                     /gene="F23E12.20"
     CDS             join(3124. .3241,3323. .3465,3595. .3696,3945. .4107,
                     4211. .4428)
                     /gene="F23E12.20"
                     /note="similarity to dihydroflavonol-4-reductase, Gerbera
                     x sp., PIR2:S35189"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAA18727.1"
                     /db_xref="GI:3080408"
                     /db_xref="SPTREMBL:O65487"
                     /translation="MDQAKGKVCVTGASGFLASWLVKRLLLEGYEVIGTVRDPGNEKK
                     LAHLWKLEGAKERLRLVKADLMEEGSFDNAIMGCQGVFHTASPEILRPAIEGTLNVLR
                     SCRKNPSLKRVVLTSSSSTVWYALSKTLAEQAAWKFSEENGIDLVTVLPSFLVGPSLP
                     PDLCSTASDVLGLLKGETEKFQWHGQMGYVHIDDVARTHIVVFEHEAAQGRYICSSNV
                     ISLEELVSFLSARYPSLPIPKRFDTDPLN"
     intron          3242. .3322
                     /gene="F23E12.20"
                     /number=1
     exon            3323. .3465
                     /gene="F23E12.20"
                     /number=2
     intron          3466. .3594
                     /gene="F23E12.20"
                     /number=2
     exon            3595. .3696
                     /gene="F23E12.20"
                     /number=3
     intron          3697. .3944
                     /gene="F23E12.20"
                     /number=3
     exon            3945. .4107
                     /gene="F23E12.20"
                     /number=4
     intron          4108. .4210
                     /gene="F23E12.20"
                     /number=4
     exon            4211. .4428
                     /gene="F23E12.20"
                     /number=5
     gene            complement(join(4939. .5013,5109. .5195,5307. .5342,
                     5451. .5651,5725. .5811,6161. .6163))
                     /gene="F23E12.30"
     gene            4939. .6163
                     /gene="F23E12.30"
     CDS             complement(join(4939. .5013,5109. .5195,5307. .5342,
                     5451. .5651,5725. .5811,6161. .6163))
                     /gene="F23E12.30"
                     /note="Contains Clathrin adaptor complexes small chain
                     signature [VVYKRYASLYF]
                     contains EST gb:T44219"
                     /codon_start=1
                     /product="clathrin assembly protein AP19 homolog"
                     /protein_id="CAA18728.1"
                     /db_xref="GI:3080409"
                     /db_xref="SPTREMBL:O23685"
                     /translation="MIHFVLLVSRQGKVRLTKWYSPYAQKERSKVIRELSGVILNRGP
                     KLCNFVEWRGYKVVYKRYASLYFCMCIDQEDNELEVLEIIHHYVEILDRYFGSVCELD
                     LIFNFHKAYYILDELLIAGELQESSKKTVARIISAQDQLVEVAKEEASSISNIIAQAT
                     NR"
     exon            complement(4939. .5013)
                     /gene="F23E12.30"
                     /number=1
     intron          complement(5014. .5108)
                     /gene="F23E12.30"
                     /number=1
     exon            complement(5109. .5195)
                     /gene="F23E12.30"
                     /number=2
     intron          complement(5196. .5306)
                     /gene="F23E12.30"
                     /number=2
     exon            complement(5307. .5342)
                     /gene="F23E12.30"
                     /number=3
     intron          complement(5343. .5450)
                     /gene="F23E12.30"
                     /number=3
     exon            complement(5451. .5651)
                     /gene="F23E12.30"
                     /number=4
     intron          complement(5652. .5724)
                     /gene="F23E12.30"
                     /number=4
     exon            complement(5725. .5811)
                     /gene="F23E12.30"
                     /number=5
     intron          complement(5812. .6160)
                     /gene="F23E12.30"
                     /number=5
```

```
     exon            complement(6161. .6163)
                     /gene="F23E12.30"
                     /number=6
     exon            complement(7352. .7582)
                     /gene="F23E12.40"
                     /number=1
     gene            7352. .7582
                     /gene="F23E12.40"
     gene            complement(7352. .7582)
                     /gene="F23E12.40"
     CDS             complement(7352. .7582)
                     /gene="F23E12.40"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAA18729.1"
                     /db_xref="GI:3080410"
                     /db_xref="SPTREMBL:O65488"
                     /translation="MPLVRKNGVVMRIGSPMLHESENLVGDGVEAPPRGERDEIGEEV
                     GGETNVVSESCGGEEGAQDAAKTTPEGEADRI"
     gene            8233. .9045
                     /gene="F23E12.50"
     gene            complement(8233. .9045)
                     /gene="F23E12.50"
     exon            complement(8233. .9045)
                     /gene="F23E12.50"
                     /number=1
     CDS             complement(8233. .9045)
                     /gene="F23E12.50"
                     /note="similarity to DNA-binding protein PD1, Pisum
                     sativum"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAA18730.1"

  Query Match             97.8%;  Score 952.8;  DB 8;  Length 86710;
  Best Local Similarity   99.7%;  Pred. No. 2.1e-258;
  Matches  965;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy       7 gacctgcctctacagagacctgaagattccagaaccccacctgatcaaaaataacatgga 66
          || |||| ||||||||||||||||||||||||||||||||||| ||||||||||||||
Db    9099 GAACTGCATCTACAGAGACCTGAAGATTCCAGAACCCCACCTGATC-AAAATAACATGGA 9041

Qy      67 acttaacagatctgaagcagacgaagcaaaggccgagaccactcccaccggtggagccac 126
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9040 ACTTAACAGATCTGAAGCAGACGAAGCAAAGGCCGAGACCACTCCCACCGGTGGAGCCAC 8981

Qy     127 cagctcagccacagcctctggctcttcctccggacgtcgtccacgtggtcgtcctgcagg 186
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8980 CAGCTCAGCCACAGCCTCTGGCTCTTCCTCCGGACGTCGTCCACGTGGTCGTCCTGCAGG 8921

Qy     187 ttccaaaaacaaacccaaacctccgacgattataactagagatagtcctaacgtccttag 246
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8920 TTCCAAAAACAAACCCAAACCTCCGACGATTATAACTAGAGATAGTCCTAACGTCCTTAG 8861

Qy     247 atcacacgttcttgaagtcacctccggttcggacatatccgaggcagtctccacctacgc 306
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8860 ATCACACGTTCTTGAAGTCACCTCCGGTTCGGACATATCCGAGGCAGTCTCCACCTACGC 8801

Qy     307 cactcgtcgcggctgcggcgtttgcattataagcggcacgggtgcggtcactaacgtcac 366
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8800 CACTCGTCGCGGCTGCGGCGTTTGCATTATAAGCGGCACGGGTGCGGTCACTAACGTCAC 8741

Qy     367 gatacggcaacctgcggctccggctggtggaggtgtgattaccctgcatggtcggtttga 426
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8740 GATACGGCAACCTGCGGCTCCGGCTGGTGGAGGTGTGATTACCCTGCATGGTCGGTTTGA 8681

Qy     427 cattttgtctttgaccggtactgcgcttccaccgcctgcaccaccgggagcaggaggttt 486
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8680 CATTTTGTCTTTGACCGGTACTGCGCTTCCACCGCCTGCACCACCGGGAGCAGGAGGTTT 8621

Qy     487 gacggtgtatctagccggaggtcaaggacaagttgtaggagggaatgtggctggttcgtt 546
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8620 GACGGTGTATCTAGCCGGAGGTCAAGGACAAGTTGTAGGAGGGAATGTGGCTGGTTCGTT 8561

Qy     547 aattgcttcgggaccggtagtgttgatggctgcttctttttgcaaacgcagtttatgatag 606
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8560 AATTGCTTCGGGACCGGTAGTGTTGATGGCTGCTTCTTTTGCAAACGCAGTTTATGATAG 8501

Qy     607 gttaccgattgaagaggaagaaaccccaccgccgagaaccaccgggggtgcagcagcagca 666
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8500 GTTACCGATTGAAGAGGAAGAAACCCCACCGCCGAGAACCACCGGGGTGCAGCAGCAGCA 8441

Qy     667 gccggaggcgtctcagtcgtcggaggttacggggagtggggcccaggcgtgtgagtcaaa 726
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8440 GCCGGAGGCGTCTCAGTCGTCGGAGGTTACGGGGAGTGGGGCCCAGGCGTGTGAGTCAAA 8381

Qy     727 cctccaaggtggaaatggtggaggaggtgttgctttctacaatcttggaatgaatatgaa 786
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8380 CCTCCAAGGTGGAAATGGTGGAGGAGGTGTTGCTTTCTACAATCTTGGAATGAATATGAA 8321

Qy     787 caattttcaattctccgggggagatatttacggtatgagcggcggtagcggaggaggtgg 846
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8320 CAATTTTCAATTCTCCGGGGGAGATATTTACGGTATGAGCGGCGGTAGCGGAGGAGGTGG 8261

Qy     847 tggcggtgcgactagacccgcgtttttagagttttagcgttttggtgacacctttttgttgc 906
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8260 TGGCGGTGCGACTAGACCCGCGTTTTAGAGTTTTAGCGTTTTGGTGACACCTTTTGTTGC 8201

Qy     907 gtttgcgtgtttgacctcaaactactaggctactagctatagcggttgcgaaatgcgaat 966
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8200 GTTTGCGTGTTTGACCTCAAACTACTAGGCTACTAGCTATAGCGGTTGCGAAATGCGAAT 8141

Qy     967 attaggtt 974
          ||||||||
Db    8140 ATTAGGTT 8133


RESULT   2
ATCHRIV83
LOCUS       ATCHRIV83              197859 bp    DNA     linear   PLN 16-MAR-2000
DEFINITION  Arabidopsis thaliana DNA chromosome 4, contig fragment No. 83.
ACCESSION   AL161587
VERSION     AL161587.2  GI:7270470
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 16641 to 103350)
  AUTHORS   Hilbert,H., Braun,M., Holzer,E., Brandt,A., Duesterhoeft,A.,
            Mewes,H.W., Lemcke,K. and Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   2  (bases 98352 to 125759)
  AUTHORS   Robben,J., Grymonprez,B., Volckaert,G., Mewes,H.W., Lemcke,K. and
            Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   3  (bases 120761 to 197859)
  AUTHORS   Rose,M., Hempel,S., Entian,K.-D., Mewes,H.W., Lemcke,K. and
            Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   4  (bases 1 to 24256)
  AUTHORS   Murphy,G., Ridley,P., Hudson,S., Mewes,H.W., Lemcke,K. and
            Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   5  (bases 1 to 197859)
  AUTHORS   EU Arabidopsis sequencing,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
COMMENT     Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
```

```
          and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
          this fragment has an overlap with ATCHRIV82 at the 5' end and an
          overlap with ATCHRIV84 at the 3' end.
FEATURES             Location/Qualifiers
     source          1. .197859
                     /organism="Arabidopsis thaliana"
                     /variety="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="4"
     gene            11339. .12705
                     /gene="AT4g35190"
     CDS             join(11339. .11494,11536. .11664,11853. .11952,12119. .12189,
                     12314. .12457,12499. .12705)
                     /gene="AT4g35190"
                     /note="similarity to ctf, Mycobacterium leprae, gb:U15180
                     contains EST gb:T45691, T22640, T75954"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAB80236.1"
                     /db_xref="GI:7270471"
                     /translation="MEIVKSRFKRVCVFCGSSSGKRECYSDAATDLAQELVRLCLNLN
                     ESLENLKWVTRRLNLVYGGGSIGLMGLVSQAVHEAGGHVLGYAQIYDLFTLITGETYG
                     EVIAVADMHERKAEMARHSDCFIALPGGYGTLEELLEVIAWAQLGIHDKPVGLLNVDG
                     YYNYLLTFIDKAVDDGFIKPSQRHIFVSAPNAKELVQKLEILMKINKDKKFDSAIDLL
                     FFDIPQAYKPVNDGVIAKSRWEVEKKVQQPQQQQQVVFCSNTSMQTEIAL"
     exon            11339. .11494
                     /gene="AT4g35190"
                     /number=1
     intron          11495. .11535
                     /gene="AT4g35190"
                     /number=1
     exon            11536. .11664
                     /gene="AT4g35190"
                     /number=2
     intron          11665. .11852
                     /gene="AT4g35190"
                     /number=2
     exon            11853. .11952
                     /gene="AT4g35190"
                     /number=3
     intron          11953. .12118
                     /gene="AT4g35190"
                     /number=3
     exon            12119. .12189
                     /gene="AT4g35190"
                     /number=4
     intron          12190. .12313
                     /gene="AT4g35190"
                     /number=4
     exon            12314. .12457
                     /gene="AT4g35190"
                     /number=5
     intron          12458. .12498
                     /gene="AT4g35190"
                     /number=5
     exon            12499. .12705
                     /gene="AT4g35190"
                     /number=6
     gene            13757. .14518
                     /gene="AT4g35200"
     gene            complement(13757. .14518)
                     /gene="AT4g35200"
     CDS             complement(13757. .14518)
                     /gene="AT4g35200"
                     /note="similarity to various predicted proteins,
                     Arabidopsis thaliana"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAB80237.1"
                     /db_xref="GI:7270472"
                     /translation="MAVSFHVRSNSYPSRQHPQAAHVDEQLTRLRSSDSASSSSICQR
                     LSNLQDLHDSLEKMIRLSVTNLALSQDQIEKLLDGSLRILDLCNIAKDAISQMKEGLM
                     EIQSILRRKPGDLSGEVKKYLVSRKFLKKSLQKVIKSLKVCQSKDSTNASLVVFGRAE
                     AVTMALFESLFSFMSGSKACGKWSLVSKMMSQNKVTCEAEANEFTRIDSEFQSEKSLQ
                     MEDVQNLESCIQDLEDGIESLSKSLIKYRVSILNI"
     exon            complement(13757. .14518)
                     /gene="AT4g35200"
                     /number=1
     gene            16043. .16795
                     /gene="AT4g35210"
     exon            16043. .16795
                     /gene="AT4g35210"
                     /number=1
     CDS             16043. .16795
                     /gene="AT4g35210"
                     /note="similarity to various predicted proteins,
                     Arabidopsis thaliana"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAB80238.1"
                     /db_xref="GI:7270473"
                     /translation="MAVSFHVRSSSYPSRQHPQAAHVDEQLTRLRSSGTASSSSICQR
                     LSNLQDLHDSLEKMIRLSVTNQALSQDQIEKLLDGSIKILDLCSISKDGLSQMKESLK
                     EIQSIVRRKRGDLSAEVKKYLASRKFLKKSFEKVLKSLKTSQNKNDALAVFGEAETVT
                     IALFESLFSFMSGSKACGKWSLVSKMMSQSKGTCEAEANEFTRVDMEFQSEKSLQMED
                     VQNLEICIQDLEDGIGSLSKSLIKYRVSILNI"
     misc_feature    16640. .24256
                     /note="position 76883-84499 overlaps to BAC clone F23E12,
                     EMBL acc:AL022604; for sequence analysis please refer to
                     this accession"
     exon            17252. .17657
                     /gene="AT4g35220"
                     /number=1
     gene            17252. .18625
                     /gene="AT4g35220"
     CDS             join(17252. .17657,17740. .17784,18534. .18625)
                     /gene="AT4g35220"
                     /note="similarity to predicted protein, Arabidopsis
                     thaliana
                     contains EST gb:F14002, T46305, T43454, AI996658.1"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAB80239.1"
                     /db_xref="GI:7270474"
                     /translation="MAVPPLFFLLTLLSLPSLLISAGASNAYPSIPGTAPIDGGFTDE
                     LKPIRREVYGNGKIYDISHRYTPEMPSWDSSEGIGRFLWLAASMKNGSLANNSEMKIP
                     THTGTHVDSPGHVYDKYYDAGFDVDSLDLQVLNGLALLVDVPKDKNITDWLEQKGLQF
                     VASSSIDSLPSQTCQIRNCP"
     intron          17658. .17739
                     /gene="AT4g35220"
                     /number=1
     exon            17740. .17784
                     /gene="AT4g35220"
                     /number=2
     intron          17785. .18533
                     /gene="AT4g35220"
                     /number=2
     exon            18534. .18625
                     /gene="AT4g35220"
                     /number=3
     gene            19941. .22657
                     /gene="AT4g35230"
     gene            complement(join(19941. .20231,20345. .20458,20546. .20710,
                     20792. .20980,21340. .21425,21551. .21656,21755. .21888,
                     22078. .22213,22304. .22657))
                     /gene="AT4g35230"
     CDS             complement(join(19941. .20231,20345. .20458,20546. .20710,
                     20792. .20980,21340. .21425,21551. .21656,21755. .21888,
                     22078. .22213,22304. .22657))
                     /gene="AT4g35230"
                     /note="similarity to protein kinase APK1, Arabidopsis
                     thaliana, PIR2:S28615
                     contains EST gb:AI997799.1, F13911"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAB80240.1"
```

```
                     /db_xref="GI:7270475"
                     /translation="MGCCQSLFSGDNPLGKDGVQPQPLSQNNHGGATTADNGGSGGAS
                     GVGGGGGGGIPSFSEFSFADLKAATNNFSSDNIVSESGEKAPNLVYKGRLQNRRWIA
                     VKKFTKMAWPEPKQFAEEAWGVGKLRHNRLANLIGYCCDGDERLLVAEFMPNDTLAKH
                     LFHWENQTIEWAMRLRVGYYIAEALDYCSTEGRPLYHDLNAYRVLFDEDGDPRLSCFG
                     LMKNSRDGKSYSTNLAYTPPEYLRNGRVTPESVTYSFGTVLLDLLSGKHIPPSHALDM
                     IRGKNIILLMDSHLEGKFSTEEATVVVELASQCLQYEPRERPNTKDLVATLAPLQTKS
                     DVPSYVMLGIKKQEEAPSTPQRPLSPLGEACSRMDLTAIHQILVMTHYRDDEGTNELS
                     FQEWTQQMKDMLDARKRGDQSFREKDFKTAIDCYSQNSYGYIHCFCGQFIDVGTMVSP
                     TVFGRRSLCYLLCDQPDAALRDAMQAQCVYPDWPTAFYMQSVALAKLNMNTDAADMLN
                     EAAQLEEKRQRGGRGS"
     exon            complement(19941. .20231)
                     /gene="AT4g35230"
                     /number=1
     intron          complement(20232. .20344)
                     /gene="AT4g35230"
                     /number=1

  Query Match             97.8%;  Score 952.8;  DB 8;  Length 197859;
  Best Local Similarity   99.7%;  Pred. No. 2.2e-258;
  Matches  965;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy       7 gacctgcctctacagagacctgaagattccagaaccccacctgatcaaaaataacatgga 66
          || |||| ||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   94252 GAACTGCATCTACAGAGACCTGAAGATTCCAGAACCCCACCTGATC-AAAATAACATGGA 94310

Qy      67 acttaacagatctgaagcagacgaagcaaaggccgagaccactcccaccggtggagccac 126
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   94311 ACTTAACAGATCTGAAGCAGACGAAGCAAAGGCCGAGACCACTCCCACCGGTGGAGCCAC 94370

Qy     127 cagctcagccacagcctctggctcttcctccggacgtcgtccacgtggtcgtcctgcagg 186
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   94371 CAGCTCAGCCACAGCCTCTGGCTCTTCCTCCGGACGTCGTCCACGTGGTCGTCCTGCAGG 94430

Qy     187 ttccaaaaacaaacccaaacctccgacgattataactagagatagtcctaacgtccttag 246
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   94431 TTCCAAAAACAAACCCAAACCTCCGACGATTATAACTAGAGATAGTCCTAACGTCCTTAG 94490

Qy     247 atcacacgttcttgaagtcacctccggttcggacatatccgaggcagtctccacctacgc 306
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   94491 ATCACACGTTCTTGAAGTCACCTCCGGTTCGGACATATCCGAGGCAGTCTCCACCTACGC 94550

Qy     307 cactcgtcgcggctgcggcgtttgcattataagcggcacgggtgcggtcactaacgtcac 366
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   94551 CACTCGTCGCGGCTGCGGCGTTTGCATTATAAGCGGCACGGGTGCGGTCACTAACGTCAC 94610

Qy     367 gatacggcaacctgcggctccggctggtggaggtgtgattaccctgcatggtcggtttga 426
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   94611 GATACGGCAACCTGCGGCTCCGGCTGGTGGAGGTGTGATTACCCTGCATGGTCGGTTTGA 94670

Qy     427 cattttgtctttgaccggtactgcgcttccaccgcctgcaccaccgggagcaggaggttt 486
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   94671 CATTTTGTCTTTGACCGGTACTGCGCTTCCACCGCCTGCACCACCGGGAGCAGGAGGTTT 94730

Qy     487 gacggtgtatctagccggaggtcaaggacaagttgtaggagggaatgtggctggttcgtt 546
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   94731 GACGGTGTATCTAGCCGGAGGTCAAGGACAAGTTGTAGGAGGGAATGTGGCTGGTTCGTT 94790

Qy     547 aattgcttcgggaccggtagtgttgatggctgcttcttttgcaaacgcagtttatgatag 606
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   94791 AATTGCTTCGGGACCGGTAGTGTTGATGGCTGCTTCTTTTGCAAACGCAGTTTATGATAG 94850

Qy     607 gttaccgattgaagaggaagaaaccccaccgccgagaaccaccggggtgcagcagcagca 666
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   94851 GTTACCGATTGAAGAGGAAGAAACCCCACCGCCGAGAACCACCGGGGTGCAGCAGCAGCA 94910

Qy     667 gccggaggcgtctcagtcgtcggaggttacggggagtggggcccaggcgtgtgagtcaaa 726
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   94911 GCCGGAGGCGTCTCAGTCGTCGGAGGTTACGGGGAGTGGGGCCCAGGCGTGTGAGTCAAA 94970

Qy     727 cctccaaggtggaaatggtggaggaggtgttgctttctacaatcttggaatgaatatgaa 786
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   94971 CCTCCAAGGTGGAAATGGTGGAGGAGGTGTTGCTTTCTACAATCTTGGAATGAATATGAA 95030

Qy     787 caatttttcaattctccgggggagatatttacggtatgagcggcggtagcggaggaggtgg 846
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   95031 CAATTTTCAATTCTCCGGGGGAGATATTTACGGTATGAGCGGCGGTAGCGGAGGAGGTGG 95090

Qy     847 tggcggtgcgactagacccgcgttttagagttttagcgttttggtgacacctttttgttgc 906
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   95091 TGGCGGTGCGACTAGACCCGCGTTTTAGAGTTTTAGCGTTTTGGTGACACCTTTTGTTGC 95150

Qy     907 gtttgcgtgtttgacctcaaactactaggctactagctatagcggttgcgaaatgcgaat 966
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   95151 GTTTGCGTGTTTGACCTCAAACTACTAGGCTACTAGCTATAGCGGTTGCGAAATGCGAAT 95210

Qy     967 attaggtt 974
           ||||||||
Db   95211 ATTAGGTT 95218


RESULT   3
AF194974
LOCUS       AF194974                1323 bp    mRNA    linear   PLN 10-NOV-1999
DEFINITION  Arabidopsis thaliana ESCAROLA (ESC) mRNA, complete cds.
ACCESSION   AF194974
VERSION     AF194974.1  GI:6319179
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1323)
  AUTHORS   Weigel,D., Ahn,J.H., Blazquez,M.A., Borevitz,J., Christensen,S.K.,
            Fankhauser,C., Ferrandiz,C., Kardailsky,I., Neff,M.M., Nguyen,J.T.,
            Sato,S., Wang,Z., Xia,Y., Dixon,R.A., Harrison,M.J., Lamb,C.J.,
            Yanofsky,M.F. and Chory,J.
  TITLE     Activation Tagging in Arabidopsis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1323)
  AUTHORS   Nguyen,J.T., Christensen,S.K. and Weigel,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-1999) Plant Biology Laboratory, The Salk
            Institute for Biological Studies, 10010 N. Torrey Pines Road, La
            Jolla, CA 92037, USA
FEATURES             Location/Qualifiers
     source          1. .1323
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
     5'UTR           1. .274
                     /gene="ESC"
     gene            1. .1323
                     /gene="ESC"
     CDS             275. .1210
                     /gene="ESC"
                     /note="overexpression causes late flowering,modified leaf
                     development; similar to F9H16.12 encoded by GenBank
                     Accession Number AC007369; similar to AT-hook proteins and
                     pea PD1 homolog"
                     /codon_start=1
                     /product="ESCAROLA"
                     /protein_id="AAF07197.1"
                     /db_xref="GI:6319180"
                     /translation="MEGGYEQGGGASRYFHNLFRPEIHHQQLQPQGGINLIDQHHHQH
                     QQHQQQQQPSDDSRESDHSNKDHHQQGRPDSDPNTSSSAPGKRPRGRPPGSKNKAKPP
                     IIVTRDSPNALRSHVLEVSPGADIVESVSTYARRRGRGVSVLGGNGTVSNVTLRQPVT
                     PGNGGGVSGGGGVVTLHGRFEILSLTGTVLPPPAPPGAGGLSIFLAGGQGQVVGGSVV
                     APLIASAPVILMAASFSNAVFERLPIEEEEEEGGGGGGGGGGPPQMQQAPSASPPSG
                     VTGQGQLGGNVGGYGFSGDPHLLGWGAGTPSRPPF"
     3'UTR           1211. .1323
                     /gene="ESC"
BASE COUNT      317 a    286 c    346 g    374 t
ORIGIN
```

```
  Query Match             21.8%;  Score 212.4;  DB 8;  Length 1323;
  Best Local Similarity   62.8%;  Pred. No. 5.6e-49;
  Matches  358;  Conservative    0;  Mismatches  191;  Indels   21;  Gaps    1;

Qy      80 gaagcagacgaagcaaaggccgagaccactcccaccggtggagccaccagctcagccaca 139
           ||| | ||| |  ||||    ||       ||      ||| || ||          | |
Db     449 GAATCTGACCATTCAAACAAAGATCATCATCAACAGGGTCGACCCGATTCAGACCCGAAT 508

Qy     140 gcctctggctcttcctccggacgtcgtccacgtggtcgtcctgcaggttccaaaaacaaa 199
           | ||  |||| |  | |||   ||||||||||| |||||  |||| || || ||||||
Db     509 ACATCAAGCTCAGCACCGGGAAAACGTCCACGTGGACGTCCACCAGGATCTAAGAACAAA 568

Qy     200 cccaaacctccgacgattataactagagatagtcctaacgtccttagatcacacgttctt 259
           |||| || ||||  ||  ||||| | ||||| || ||||  |||||||| |||||||||
Db     569 GCCAAGCCACCGATCATAGTAACTCGTGATAGCCCCAACGCGCTTAGATCTCACGTTCTT 628

Qy     260 gaagtcacctccggttcggacatatccgaggcagtctccacctacgccactcgtcgcggc 319
           |||||  |  | ||  | |||||   |||   || ||||| ||||| |   |  | ||
Db     629 GAAGTATCTCCTGGAGCTGACATAGTTGAGAGTGTTTCCACGTACGCTAGGAGGAGAGGG 688

Qy     320 tgcggcgtttgcattataagcggcacgggtgcggtcactaacgtcacgatacggcaacct 379
           | ||||| | | || || | || |  ||  | ||  |||||||||| | || || ||
Db     689 AGAGGCGTCTCCGTTTTAGGAGGAAACGGCACCGTATCTAACGTCACTCTCCGTCAGCCA 748

Qy     380 gcggctc--------------------cggctggtggaggtgtgattaccctgcatggt 418
           |   |||                    | |  || ||||| ||  | ||  | |||||
Db     749 GTCACTCCTGGAAATGGCGGTGGTGTGTCCGGAGGAGGAGGAGTTGTGACTTTACATGGA 808

Qy     419 cggtttgacattttgtctttgaccggtactgcgcttccaccgcctgcaccaccgggagca 478
           ||||||| ||| | ||  | || || |||   | ||||| |||||||| || || ||
Db     809 AGGTTTGAGATTCTTTCGCTAACGGGGACTGTTTTGCCACCTCCTGCACCGCCTGGTGCC 868

Qy     479 ggaggtttgacggtgtatctagccggaggtcaaggacaagttgtaggagggaatgtggct 538
           || |||||| |  | | | |||||||||| ||||| || || || ||||| |  || |
Db     869 GGTGGTTTGTCTATATTTTTAGCCGGAGGGCAAGGTCAGGTGGTCGGAGGAAGCGTTGTG 928

Qy     539 ggttcgttaattgcttcgggaccggtagtgttgatggctgcttcttttgcaaacgcagtt 598
           | | |  | ||||| || |  |||||  |  | ||||| ||||| ||  |||| || |||
Db     929 GCTCCCCTTATTGCATCAGCTCCGGTTATACTAATGGCGGCTTCGTTCTCAAATGCGGTT 988

Qy     599 tatgataggttaccgattgaagaggaagaa 628
           |  || ||  |||||||||| ||||| |||
Db     989 TTCGAGAGACTACCGATTGAGGAGGAGGAA 1018
```

```
RESULT   4
AC007369/c
LOCUS       AC007369               103192 bp    DNA     linear   PLN 17-MAY-1999
DEFINITION  Arabidopsis thaliana chromosome I BAC F9H16 genomic sequence,
            complete sequence.
ACCESSION   AC007369
VERSION     AC007369.2  GI:4757678
KEYWORDS    HTG.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 103192)
  AUTHORS   Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
            Altafi,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P.,
            Gonzalez,A., Kremenetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S.,
            Luros,S., Schwartz,J., Shinn,P., Toriumi,M., Vysotskaia,V.S.,
            Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 103192)
  AUTHORS   Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
            Altafi,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P.,
            Gonzalez,A., Kremenetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S.,
            Luros,S., Schwartz,J., Shinn,P., Toriumi,M., Vyotskaia,V.,
            Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-APR-1999) DNA Sequencing and Technology Center,
            Stanford University, 855 California Avenue, Palo Alto, CA 94304,
            USA
REFERENCE   3  (bases 1 to 103192)
  AUTHORS   Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
            Altafi,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P.,
            Gonzalez,A., Kremenetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S.,
            Luros,S., Schwartz,J., Shinn,P., Toriumi,M., Vyotskaia,V.,
            Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAY-1999) DNA Sequencing and Technology Center,
            Stanford University, 855 California Avenue, Palo Alto, CA 94304,
            USA
REFERENCE   4  (bases 1 to 103192)
  AUTHORS   Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
            Altafi,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P.,
            Gonzalez,A., Kremenetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S.,
            Luros,S., Schwartz,J., Shinn,P., Toriumi,M., Vyotskaia,V.,
            Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-1999) DNA Sequencing and Technology Center,
            Stanford University, 855 California Avenue, Palo Alto, CA 94304,
            USA
COMMENT     On May 7, 1999 this sequence version replaced gi:4678189.
            e-mail for correspondence: arab@sequence.stanford.edu
            Genes with similarity to proteins in the databases are described as
            'putative', '-like' or 'similar to'.  Genes that have EST
            similarity but no significant protein similarity are described as
            'unknown proteins'.  Genes that are annotated based only on gene
            prediction software are described as 'hypothetical proteins'.
            The software programs used to predict genes include:  Grail
            (Informatics Group, Oak Ridge National Laboratory,
            http://compbio.ornl.gov/section/index.html), GENSCAN (Chris Burge,
            http://gnomic.stanford.edu/~chris/GENSCANW.html), Fexa (V.Solovyev
            & A.Salamov, Sanger Centre,  http://genomic.sanger.ac.uk/), and
            NetPlantGene (S.M. Hebsgaard, et al., CBS, Technical University of
            Denmark,http://www.cbs.dtu.dk/NetPlantGene.html).
FEATURES              Location/Qualifiers
     source           1. .103192
                      /organism="Arabidopsis thaliana"
                      /cultivar="Columbia"
                      /db_xref="taxon:3702"
                      /chromosome="I"
                      /map="mi203"
                      /clone="F9H16"
     gene             complement(8784. .9859)
                      /gene="F9H16.1"
     CDS              complement(join(8784. .9101,9359. .9574,9662. .9859))
                      /gene="F9H16.1"
                      /note="Unknown protein; Location of ESTs 152C14T7 ,
                      gb|T88012, 173P7T7 , gb|H36502 and 107E18T7 , gb|T22775"
                      /codon_start=1
                      /protein_id="AAD30591.1"
                      /db_xref="GI:4836888"
                      /translation="MIRTEEEEDYTSPPWLMPMLRGSYFVPCSIHVDSNKNECNLFCL
                      DCAGNAFCSYCLVKHKDHRVVQIRRSSYHNVVRVNEIQKFIDIACVQTYIINSAKIVF
                      LNERPQPRIGKGVTNTCEICCRSLLDSFRFCSLGCKLGGMRRGDLSLTFSLKGKHGRE
                      YLGGSESDEATTPTKMRKTNAFNRLMSGLSISTVRFDDYGPNGDQRSSSSGDEGGFSF
                      SPGTPPIYNHRNSSRRKGVPHRAPF"
     gene             15929. .17348
                      /gene="F9H16.2"
     CDS              join(15929. .16340,16804. .17348)
                      /gene="F9H16.2"
                      /note="Hypothetical protein"
                      /codon_start=1
                      /protein_id="AAD30592.1"
                      /db_xref="GI:4836889"
                      /translation="MRTKSIVINSNSKNDTSTISATTALRRSASNTTPSAAVVRTNDL
                      NKKPRPRSVEFPASPQLVQGEEMVHFGHPQHVLVKVELPDIYTCAGCKEEGAGVRYVC
                      QECDYQLHEFCALAPPQLKSHPFHYQHQLLFFAKPAKGGIVKSKCDVCGRSPKGYTFR
                      CKACSFQMHPGCAMLSPSLSSSSLHHHPLRLLPSSSAGSTTGGDSGGFLCGECKRGKR
                      TGRVYRCTVCDYHLHAVCAKDAAVNGLRANGHKGRDKSPAVLGTAARLASQVVIDFLG
                      GIIDGLGEGVGEAIIDGVTRGGGGGGGNGGGTRVIPRVRGG"
     gene             complement(18939. .22830)
```

```
                     /gene="F9H16.3"
     CDS             complement(join(18939. .19358,19441. .20040,20119. .20263,
                     20346. .20481,20575. .20692,20776. .20883,20975. .21049,
                     21141. .21943,22041. .22246,22334. .22830))
                     /gene="F9H16.3"
                     /note="Unknown protein; location of ESTs: 111G20XP 3',
                     gb|AA585893, FAI314 5' end, gb|Z37590 and 111G20T7,
                     gb|T42259"
                     /codon_start=1
                     /protein_id="AAD30593.1"
                     /db_xref="GI:4836890"
                     /translation="MDEVGAQVAAPMFIHQSLGRKRDLYYPMSNRLVQSQPQRRDEWN
                     SKMWDWDSRRFEAKPVDVEVQEFDLTLRNRSGEERGLDLNLGSGLTAVEETTTTQNV
                     RPNKKVRSGSPGGNYPMCQVDNCTEDLSHAKDYHRRHKVCEVHSKATKALVGKQMQRF
                     CQQCSRFHLLSEFDEGKRSCRRRLAGHNRRRRKTTQPEEVASGVVVPGNHDTTNNTAN
                     ANMDLMALLTALACAQGKNAVKPPVGSPAVPDREQLLQILNKINALPLPMDLVSKLNN
                     IGSLARKNMDHPTVNPQNDMNGASPSTMDLLAVLSTTLGSSSPDALAILSQGGFGNKD
                     SEKTKLSSYENGVTTNLEKRTFGFSSVGGERSSSSNQSPSQDSDSRGQDTRSSLSLQL
                     FTSSPEDESRPTVASSRKYYSSASSNPVEDRSPSSSPVMQELFPLQASPETMRSKNHK
                     NSSPRTGCLPLELFGASNRGAADPNFKGFGQQSGYASSGSDYSPPSLNSDAQDRTGKI
                     VFKLLDKDPSQLPGTLRSEIYNWLSNIPSEMESYIRPGCVVLSVYVAMSPAAWEQLEQ
                     KLLQRLGVLLQNSPSDFWRNARFIVNTGRQLASHKNGKVRCSKSWRTWNSPELISVSP
                     VAVVAGEETSLVVRGRSLTNDGISIRCTHMGSYMAMEVTRAVCRQTIFDELNVNSFKV
                     QNVHPGFLGRCFIEVENGFRGDSFPLIIANASICKELNRLGEEFHPKSQDMTEEQAQS
                     SNRGPTSREEVLCFLNELGWLFQKNQTSELREQSDFSLARFKFLLVCSVERDYCALIR
                     TLLDMLVERNLVNDELNREALDMLAEIQLLNRAVKRKSTKMVELLIHYLVNPLTLSSS
                     RKFVFLPNITGPGGITPLHLAACTSGSDDMIDLLTNDPQEIGLSSWNTLRDATGQTPY
                     SYAAIRNNHNYNSLVARKLADKRNKQVSLNIEHEVVDQTGLSKRLSLEMNKSSSSCAS
                     CATVALKYQRRVSGSQRLFPTPIIHSMLAVATVCVCVCVFMHAFPIVRQGSHFSWGGL
                     DYGSI"
     gene            complement(27992. .33534)
                     /gene="F9H16.4"
     CDS             complement(join(27992. .28280,28670. .28985,29062. .29743,
                     29822. .29920,30018. .30212,30431. .30610,30713. .30871,
                     30713. .30871,30958. .33534))
                     /gene="F9H16.4"
                     /note="Hypothetical protein"
                     /codon_start=1
                     /protein_id="AAD30594.1"
                     /db_xref="GI:4836891"
                     /translation="MPVEVERDQGEVSVKVDFENATEIKPEVVVSATKEDVVNGISHG
                     GSNNGNGNDTDGSYDFITENDTVGDDFVESDYVKPVDDANVEKDLKEGENVKVDAPSI
                     ADDDVLGVSQDSQTLEKSELESTDDGPEEVVEIPKSEVEDSLEKSVDQQHPGNGHLES
                     GLEGKVESKEEVEQLHDSEVGSKDLTKNNVEEPEVEIESDSETDVEGHQGDKIEAQEK
                     SDRDLDVSQDLKLNENVEKHPVDSDEVRESELVSAKVSPTEPSDGGMDLGQPTVTDPA
                     ETINGSESVNDHVGSEPVTVLEPVSVENGHPPVESELERSSDVPFTSVAEKVNASDGE
                     VLPDSGTVDVVVSEVSSDVPAETQALNAISLDSQPSGKDSVVENGNSKSESEDSKMQS
                     EIGAVDDGSVSDGSINTHPESQDASDPTCDQGGKQHISSEVKEVLDAPASEEISDAVI
                     VAKDNGSEAAISDGLSCTNQQGSESDEISGLVEKLPSHALHEVVSSANDTSVIVSDDT
                     KSQGLSEDHGVDTNQTIQDDCSAELEEVTDVNVKHAPNEKVQGDNSEGNLNVGGDVCL
                     NSAEEAKELPTGDLSGNASHESAETLSTNIDEPLSLLDTKTAVSDFAESSAGVAGEID
                     AVAMESEAAQSIKQCAEAHVAPSIIEDGEIDREVNCGSEVNVTKTTPVAVREDIPPKE
                     VSEMEESDVKERSSINTDEEVATASVASEIKTCAQDLESKVVTSTDTIHTGAKDCVDS
                     QPAENKEGNKLIKNEIRLCTSLVENQKDGVDSIYKLLCSGNVVDRTDDKVASTGEVSV
                     LDASEGLTVAAEIEKRPFYFLPRVPRYDDEKLAEQLKHAEEQVDQKTQNRDALRADIQ
                     KIRAICKDYDISYKAVMAEERSARKAMHSKRQEIEALQSMISRVKSAASVDDIDSRAI
                     CKDYDISYKAVMAEERSARKAMHSKRQEIEALQSMISRVKSAASVDDIDSRVRNMEHT
                     MQHTTLSLNEEKGFMREIKQLKQLREQISSSMGTKDEVKQALDEKEKTEERLKVLRKE
                     LDALRNDLSKAEEITKAAKKKCDGEWEAQSKLQEQFRAADAVRQEAFVHLQDLKKQQR
                     EKNKYFFKYRDNSRAASEMALKKDRAALQSLCSDQVENFMNMWNNDDEFRKYYVKSNT
                     RSTFRRLGTLDGRSLGPDEEPPRITYAPRTDKLRTSSDRAEKHEAVPAQKEKVVKFEG
                     SKVENNGKEVAKPTEQKSQTTKSKKAVKPDQPPSIVTELVSGKEEIEKSATPEEEEPP
                     KLTKEEEELIKKEEEKRKQKEAAKMKEQHRLEEIAKAKEAMERKKKREEKAKARAVLK
                     AQKEAEEREKVKAYFCHSVIQIGFRSVSLARTREEAKEEGEKKGDFYIRRDSNRKPNS
                     DSRNCSRNPKGDRNSKETNRRGESTNEEISQTFITVSETKQVKIGSSAFKKPRKQEKT
                     AAMDVDWTHSCDHHRIVPSRSSTIVRSCHISHPGVVPHQCQEIVHLDFPSLFKDRAEK
                     HETVPPVSALDSVLYLPLHQNHLRDQILEQSMRRTRTASRSSATFCLDNRSSLTRTP"
     gene            37958. .45281
                     /gene="F9H16.5"
     CDS             join(37958. .40897,41604. .41993,42096. .45281)
                     /gene="F9H16.5"
                     /note="Putative RNA helicase; Highly similar to human U5
                     snRNP-specific 200kD protein, gi|3255965 and yeast
                     pre-mRNA splicing helicase BRR2, gi|731380 and other RNA
                     helicases.
                     "
                     /codon_start=1
                     /protein_id="AAD30595.1"
                     /db_xref="GI:4836892"
                     /translation="MANLGGGAEAHARFKQYEYRANSSLVLTTDNRPRDTHEPTGEPE
                     TLWGKIDPRSFGDRVAKGRPQELEDKLKKSKKKERDVVDDMVNIRQSKRRRLREESVL
                     TDTDDAVYQPKTKETRAAYEAMLGLIQKQLGGQPPSIVSGAADEILAVLKNDAFRNPE
                     KKMEIEKLLNKIENHEFDQLVSIGKLITDFQEGGDSGGGRANDDEGLDDDLGVAVEFE
                     ENEEDDEESDPDMVEEDDDEEDDEPTRTGGMQVDAGINDEDAGDANEGTNLNVQDIDA
                     YWLQRKISQAYEQQIDPQQCQVLAEELLKILAEGDDRVVEDKLLMHLQYEKFSLVKFL
                     LRNRLKVVWCTRLARAEDQEERNRIEEEMRGLGPELTAIVEQLHATRATAKEREENLQ
                     KSINEEARRLKDETGGDGGRGRRDVADRDSESGWVKGQRQMLDLESLAFDQGGLLMAN
                     KKCDLPPGSYRSHGKGYDEVHVPWVSKKVDRNEKLVKITEMPDWAQPAFKGMQQLNRV
                     QSKVYDTALFKAENILLCAPTGAGKTNVAMLTILQQLEMNRNTDGTYNHGDYKIVYVA
                     PMKALVAEVVGNLSNRLKDYGVIVRELSGDQSLTGREIEETQIIVTTPEKWDIITRKS
                     GDRTYTQLVRLLIIDEIHLLHDNRGPVLESIVARTLRQIETTKENIRLVGLSATLPNY
                     EDVALFLRVDLKKGLFKFDRSYRPVPLHQQYIGISVKKPLQRFQLMNDLCYQKVLAGA
                     GKHQVLIFVHSRKETSKTARAIRDTAMANDTLSRFLKEDSVTRDVLHSHEDIVKNSDL
                     KDILPYGFAIHHAGLSRGDREIVETLFSQGHVQVLVSTATLAWGVNLPAHTVIIKGTQ
                     VYNPEKGAWMELSPLDVMQMLGRAGRPQYDQHGEGIIITGYSELQYYLSLMNEQLPIE
                     SQFISKLADQLNAEIVLGTVQNAREACHWLGYTYLYIRMVRNPTLYGLAPDALAKDVV
                     LEERRADLIHSAATILDKNNLVKYDRKSGYFQVTDLGRIASYYYITHGTIATYNEHLK
                     PTMGDIDLYRLFSLSDEFKYVTVRQDEKMELAKLLDRVPIPIKETLEEPSAKINVLLQ
                     AYISQLKLEGLSLTSDMVYITQSAGRLVRALYEIVLKRGWAQLAEKALNLSKMVGKRM
                     WSVQTPLRQFHGLSNDILMQLEKKDLVWERYYDLSAQELGELIRSPKMGKPLHKFIHQ
                     FPKVTLSAHVQPITRTVLNVELTVTPDFLWDEKIHKYVEPFWIIVEDNDGEKILHHEY
```

```
  Query Match             21.8%;  Score 212.4;  DB 8;  Length 103192;
  Best Local Similarity   62.8%;  Pred. No. 8.6e-49;
  Matches  358;  Conservative    0;  Mismatches  191;  Indels   21;  Gaps    1;

Qy      80 gaagcagacgaagcaaaggccgagaccactcccaccggtggagccaccagctcagccaca 139
           ||| | ||| |  ||||    ||       ||      ||| || ||          | |
Db   74674 GAATCTGACCATTCAAACAAAGATCATCATCAACAGGGTCGACCCGATTCAGACCCGAAT 74615

Qy     140 gcctctggctcttcctccggacgtcgtccacgtggtcgtcctgcaggttccaaaaacaaa 199
           | ||  |||| |  | |||   ||||||||||| |||||  |||| || || ||||||
Db   74614 ACATCAAGCTCAGCACCGGGAAAACGTCCACGTGGACGTCCACCAGGATCTAAGAACAAA 74555

Qy     200 cccaaacctccgacgattataactagagatagtcctaacgtccttagatcacacgttctt 259
           |||| || |||  ||  |||||| | ||||| || ||||  |||||||| |||||||||
Db   74554 GCCAAGCCACCGATCATAGTAACTCGTGATAGCCCCAACGCGCTTAGATCTCACGTTCTT 74495

Qy     260 gaagtcacctccggttcggacatatccgaggcagtctccacctacgccactcgtcgcggc 319
           ||||| |  | ||  | ||||||   |||   || ||||| |||| |   |  | ||
Db   74494 GAAGTATCTCCTGGAGCTGACATAGTTGAGAGTGTTTCCACGTACGCTAGGAGGAGAGGG 74435

Qy     320 tgcggcgtttgcattataagcggcacgggtgcggtcactaacgtcacgatacggcaacct 379
           | ||||| | | || || | || |  ||  | ||  |||||||||||  | || || ||
Db   74434 AGAGGCGTCTCCGTTTTAGGAGGAAACGGCACCGTATCTAACGTCACTCTCCGTCAGCCA 74375

Qy     380 gcggctc----------------------cggctggtggaggtgtgattaccctgcatggt 418
           |   |||                      | |  || ||||| ||  | ||  | |||||
Db   74374 GTCACTCCTGGAAATGGCGGTGGTGTGTCCGGAGGAGGAGGAGTTGTGACTTTACATGGA 74315

Qy     419 cggtttgacattttgtctttgaccggtactgcgcttccaccgcctgcaccaccgggagca 478
           ||||||| ||| | ||  | || || ||||   | ||||| |||||||| || || ||
Db   74314 AGGTTTGAGATTCTTTCGCTAACGGGGACTGTTTTGCCACCTCCTGCACCGCCTGGTGCC 74255

Qy     479 ggaggtttgacggtgtatctagccggaggtcaaggacaagttgtaggagggaatgtggct 538
           || |||||| |  | | | |||||||||| ||||| || || || ||||| |  || |
Db   74254 GGTGGTTTGTCTATATTTTTAGCCGGAGGGCAAGGTCAGGTGGTCGGAGGAAGCGTTGTG 74195

Qy     539 ggttcgttaattgcttcgggaccggtagtgttgatggctgcttcttttgcaaacgcagtt 598
           | | |  | ||||| || |  ||||   |  | ||||| |||| ||  |||| || |||
Db   74194 GCTCCCCTTATTGCATCAGCTCCGGTTATACTAATGGCGGCTTCGTTCTCAAATGCGGTT 74135

Qy     599 tatgataggttaccgattgaagaggaagaa 628
           |  || ||  |||||||||| ||||| |||
Db   74134 TTCGAGAGACTACCGATTGAGGAGGAGGAA 74105
```

RESULT 5
AC004667
LOCUS AC004667 88411 bp DNA linear PLN 05-APR-2000
DEFINITION Arabidopsis thaliana chromosome II section 193 of 255 of the complete sequence. Sequence from clones T4C15, T32F12.
ACCESSION AC004667 AE002093
VERSION AC004667.2 GI:6598430
KEYWORDS HTG.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 88411)
AUTHORS Lin,X., Kaul,S., Rounsley,S.D., Shea,T.P., Benito,M.-I., Town,C.D., Fujii,C.Y., Mason,T.M., Bowman,C.L., Barnstead,M.E., Feldblyum,T.V., Buell,C.R., Ketchum,K.A., Lee,J.J., Ronning,C.M., Koo,H., Moffat,K.S., Cronin,L.A., Shen,M., VanAken,S.E., Umayam,L., Tallon,L.J., Gill,J.E., Adams,M.D., Carrera,A.J., Creasy,T.H., Goodman,H.M., Somerville,C.R., Copenhaver,G.P., Preuss,D., Nierman,W.C., White,O., Eisen,J.A., Salzberg,S.L., Fraser,C.M. and Venter,J.C.
TITLE Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana
JOURNAL Nature 402 (6763), 761-768 (1999)
MEDLINE 20083487
PUBMED 10617197
REFERENCE 2 (bases 1 to 88411)
AUTHORS Lin,X.
TITLE Direct Submission
JOURNAL Submitted (09-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA
COMMENT On Dec 17, 1999 this sequence version replaced gi:3668073.
The sequence and annotation of chromosome 2 were merged from those of the individual clones on this chromosome after removing overlaps. For detailed information, please see the TIGR web site (http://www.tigr.org/tdb/at/at.html).

Genes were identified by a combination of three methods: Gene prediction programs including GRAIL (ftp://arthur.epm.ornl.gov/pub/xgrail), Genefinder (Phil Green, University of Washington), Genscan (Chris Burge, http://gnomic.stanford.edu/GENSCANW.html), and NetPlantGene (http://www.cbs.dtu.dk/services/NetGene2/), searches of the complete sequence against a peptide database and plant EST databases at TIGR, and manual curations based on those analyses. Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by two or more gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats were identified by repeatmasker (Arian Smit, http://ftp.genome.washington.edu/RM/RepeatMasker.html). Genes are numbered from the top to bottom of the chromosome.

We thank the CSHL/WashU/ABI consortium for sequencing BAC clones F6P23, F5J6, T17A5, and T13L16, the ESSA group for sequencing clone F13D4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards and Satoshi Tabata for helpful assistance. In addition, we would like to thank the TIGR Bioinformatics Department, especially Lixin Zhou, Hanif Khalak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy Peterson, Michael Holmes, and Delwood Richardson for software and database support.

This work was supported by the National Science Foundation, Department of Energy and the US Department of Agriculture.

Address all correspondence to: at@tigr.org.

FEATURES Location/Qualifiers

source	1. .88411 /organism="Arabidopsis thaliana" /cultivar="Columbia" /db_xref="taxon:3702" /chromosome="II"
misc_feature	complement(<1. .88210) /note="Sequence from clone T4C15"
repeat_region	2372. .2441 /rpt_family="(TAAA)n"
mRNA	join(<2768. .2959,3319. .3416,3488. .3524,3636. .3732, 4160. .4262,5931. .6022,6392. .6444,6497. .6656,6766. .7026, 7524. .7603,7800. .>7997) /gene="At2g35070"
gene	<2768. .>7997 /gene="At2g35070"
CDS	join(2768. .2959,3319. .3416,3488. .3524,3636. .3732, 4160. .4262,5931. .6022,6392. .6444,6497. .6656,6766. .7026, 7524. .7603,7800. .7997) /gene="At2g35070" /note="hypothetical protein" /codon_start=1 /protein_id="AAC61828.1" /db_xref="GI:3668096" /translation="MAMLKRTISIVRIGRGLSYRSLSSGNHVDKLGMVKALEAINVPS TQAEALTGAITSGFESVMGKEYKSTRVAEEFRKMRADIEKMRADIATIQAEERCDHDN HIHKVEQGFNELKDEFDATKYKIVKYAIYTLIPVAAAETAASCRYMLTSFRNTTNTVH LGAVHMRIQYQTCIDKMGINEFSSFGRRGSGFQEHNGESPPFPWIPEYKVQVAVVDLW SFVQREKSMVRACNLAQGHAISLDLETAALKMSADLNRTEIPSSVSSKALLYESLKPS ISFLVKSSNLEGFTLTEDEAESVESVVMLLQKDDFKDLKTLGKMETIIQLLQTIPNQE HCSSKWCNRRVAIQCVDFSKQTIIRFGSMRRNFHDVAAIVSSSLFSSRVKEIRIRLQE HHDLLLAMLPAVVFVAEVRLSVVPEATQAHRNSSYTEVSFLELATSVKASPLLGRHLL CKEHYF"
mRNA	complement(join(<8495. .8749,9119. .9154,9181. .9248, 9563. .9584,10086. .10390,10656. .>10698)) /gene="At2g35080"
gene	complement(<8495. .>10698) /gene="At2g35080"
CDS	complement(join(8495. .8749,9119. .9154,9181. .9248, 9563. .9584,10086. .10390,10656. .10698)) /gene="At2g35080" /note="hypothetical protein" /codon_start=1 /protein_id="AAC61827.1" /db_xref="GI:3668095" /translation="MVQANTYDARNELKEITEDFLLLDLTETEADVILGYSWLSKLGE TWQDHTFAFSHNQDWISFSNKEEELKQNTNKMKMWSEYVKKGIEDEQISDCGMVVDDY LEDKFGFSGEGNGRHFPGKFIITFSRKISEKTKSVDLNHPPFSLPNIRTQFGVSLQFV SWKVAEMDKAKFEFAKRVCSFMISFVIAGAGVTTGLQMLKENQIEVEGRTEEVKEAKA RNQRLRKVKINEQWIKLQEKLFKK"
repeat_region	8809. .8842 /rpt_family="(GAAA)n"
repeat_region	9732. .9767 /rpt_family="(TAAAA)n"
repeat_region	complement(10971. .11087) /rpt_family="(CAAAA)n"
repeat_region	complement(11038. .11085) /rpt_family="POLY_A"
mRNA	join(<11437. .11628,11808. .11867,12158. .>12280) /gene="At2g35090"
gene	<11437. .>12280 /gene="At2g35090"
CDS	/note="T4C15.24; predicted by genscan" join(11437. .11628,11808. .11867,12158. .12280) /gene="At2g35090" /note="hypothetical protein" /codon_start=1 /protein_id="AAC61826.1" /db_xref="GI:3668094" /translation="MAMLKRTISIVRTGRGLSYRSLSSGNHVDKLGMVKALEAINVPS TQAEALTGAITSGLESVVEKEYQSTSVAEEVRKMRADINKGFNELKAEFETTKYEAVK

```
                     YAAYTLVAVAAVGSSMLFVMIW"
     repeat_region   12104. .12152
                     /rpt_family="(TA)n"
     mRNA            complement(join(12901. .13309,13516. .13881,14026. .14849))
                     /gene="At2g35100"
     gene            complement(12901. .14849)
                     /gene="At2g35100"
                     /note="T4C15.23"
     CDS             complement(join(13082. .13309,13516. .13881,14026. .14814))
                     /gene="At2g35100"
                     /note="unknown protein"
                     /codon_start=1
                     /protein_id="AAC61825.1"
                     /db_xref="GI:3668093"
                     /translation="MARKSSLLKRAAIAVVSVIAIYVILNASVSRSLPSSSDLPRQLI
                     REDDDDEGRAPIQPRVRVYMYNLPKRFTYGLIEQHSIARGGIKKPVGDVTTLKYPGHQ
                     HMHEWYLFSDLNQPEVDRSGSPIVRVSDPADADLFYVPVFSSLSLIVNAGRPVEAGSG
                     YSDEKMQEGLVEWLEGQEWWRRNAGRDHVIPAGDPNALYRILDRVKNAVLLVSDFGRL
                     RPDQGSFVKDVVIPYSHRVNLFNGEIGVEDRNTLLFFMGNRYRKDGGKVRDLLFQVLE
                     KEDDVTIKHGTQSRENRRAATKGMHTSKFCLNPAGDTPSACRLFDSIVSLCVPLIVSD
                     SIELPFEDVIDYRKFSIFVEANAALQPGFLVQMLRKIKTKKILEYQREMKSKFDRTKL
                     LMSMHKPNRSYHFHLNNLFVFWERPFGLWRSLRRREFTETYTSRYSPLPWHNYFSGKQ
                     TESTCMVTDS"
     mRNA            complement(join(15067. .15745,15828. .17044,17263. .17359,
                     17457. .17596,17686. .17811,18202. .18304,18567. .18712,
                     18931. .19124,19273. .19351,19461. .19571,19810. .19942,
                     20057. .20331,20626. .20733,20868. .20993,21457. .21537,
                     21745. .21825,22206. .22280,22374. .>22775))
                     /gene="At2g35110"
     gene            complement(15067. .>22775)
                     /gene="At2g35110"
                     /note="T4C15.22"
     CDS             complement(join(15220. .15745,15828. .17044,17263. .17359,
                     17457. .17596,17686. .17811,18202. .18304,18567. .18712,
                     18931. .19124,19273. .19351,19461. .19571,19810. .19942,
                     20057. .20331,20626. .20733,20868. .20993,21457. .21537,
                     21745. .21825,22206. .22280,22374. .22775))
                     /gene="At2g35110"
                     /note="unknown protein"
                     /codon_start=1
                     /protein_id="AAC61824.1"
                     /db_xref="GI:3668092"
                     /translation="MANSRQYYPSQDESMSPTSVRSREWEGPSRWTEYLGPEMAASVS
                     STRSSKQIDGHVGGSTKALNIQWVVQMIEVADGLMAKMYRLNQILEYPDPVGHVFSEA
                     FWKAGVFPNHPRICTLLSKKFPEHFSKLQLERIDKFSLDSLHDGAELHLQSLEPWIQL
                     LLDLMAFREQALRLILDLSSTVITLLPHQNSLILHAFMDLFCAFVRVNLFAEKVLEAV
                     GPSIFLSADTRKLRNEGFLSPYHPRFPDILTNSAHPMRAQDLANVTSYREWVLLGYLV
                     CPDELLRVTSIDIALVVLKENLVVTLFRDEVSLYQMVCEKEFGIGISFASADSINLTM
                     QYILLHEDYQLYVLPRVLESKKMAKSGRTKQKEADLEYSVAKQVEKMISEVHEQALQL
                     CDTIHRERRILLKQEIGRMVLFFTDQPSLLAPNIQMVFSALALAQSEVLWYFQHAGIA
                     SSRSKAARVIPVDIDPNDPTIGFLLDGMDRLCCLVRKYISAARGYALSYLSSSAGRIR
                     YLMGTPGIVALDLDPTLKGLFQRIVQHLESIPKAQGENVSAITCDLSDFRKDWLSILM
                     IVTSSRSSINIRHLEKATVSTGKEGLLSEGNAAYNWSRNTMFGPEGRPQHCCAWLSVA
                     SSFPECASLIIPEEVTKFGRDAVLYVESLIESIMGGLEGLINILDSEGGFGALESQLL
                     PEQAAAYLNNASRISAPSVKSPRVVGGFTLPGHESYPENNKSIKMLEAAIQRLTNLCS
                     ILNDMEPICVINHVFVLREYMRECILGNFKRRFLTALQTDNDLQRPSVLESLIRRHMG
                     IVHLAEQHVSMDLTQGIREILLTEAFSGPVSSLHTFEKPAEQQQTTGSAVEVVCNWYM

  Query Match             21.4%;  Score 208.6;  DB 8;  Length 88411;
  Best Local Similarity   65.5%;  Pred. No. 1e-47;
  Matches  323;  Conservative    0;  Mismatches  164;  Indels    6;  Gaps    1;

Qy     161 cgtcgtccacgtggtcgtcctgcaggttccaaaaacaaacccaaacctccgacgattata 220
           ||||| |||||||| ||||| || || || || |||||||| || || ||| |||| |
Db   76502 CGTCGGCCACGTGGCCGTCCAGCTGGATCGAAGAACAAACCGAAGCCGCCGGTGATTGTC 76561

Qy     221 actagagatagtcctaacgtccttagatcacacgttcttgaagtcacctccggttcggac 280
           ||  | || ||  | |||   |||||  | ||| ||||||||||       || |  |||
Db   76562 ACGCGCGAGAGCGCAAACACTCTTAGGGCTCACATTCTTGAAGTTGGAAGTGGCTGCGAC 76621

Qy     281 atatccgaggcagtctccacctacgccactcgtcgcggctgcggcgtttgcattataagc 340
           | | |||     ||||||| |||||   |||  |     ||||  ||||| || ||  |
Db   76622 GTTTTCGAATGTATCTCCACTTACGCTCGTCGGAGACAGCGCGGGATTTGCGTTTTATCC 76681

Qy     341 ggcacgggtgcggtcactaacgtcacgatacggcaacctgcggctccggctggtggaggt 400
           || |||||  | ||||||||||||||  || || ||      | || |||| |  |||| |
Db   76682 GGGACGGGAACCGTCACTAACGTCAGCATCCGTCA------GCCTACGGCGGCCGGAGCT 76735

Qy     401 gtgattaccctgcatggtcggtttgacattttgtctttgaccggtactgcgcttccaccg 460
           ||  | || ||||  |||   ||||| ||| | ||  |  ||||  ||   ||||| ||
Db   76736 GTTGTGACTCTGCGGGGTACTTTTGAGATTCTTTCCCTCTCCGGATCTTTTCTTCCGCCA 76795

Qy     461 cctgcaccaccgggagcaggaggtttgacggtgtatctagccggaggtcaaggacaagtt 520
           ||||| || || || ||     | |||||| | |  || || |||| |||||||||| ||
Db   76796 CCTGCTCCTCCAGGGGCGACTAGCTTGACGATATTCCTCGCTGGAGCTCAAGGACAGGTC 76855

Qy     521 gtaggagggaatgtggctggttcgttaattgcttcgggaccggtagtgttgatggctgct 580
           || ||||| || || | ||||  |||||| ||  |||| |||||| || | ||||| ||
Db   76856 GTCGGAGGTAACGTAGTTGGTGAGTTAATGGCGGCGGGGCCGGTAATGGTCATGGCAGCG 76915

Qy     581 tcttttgcaaacgcagtttatgataggttaccgattgaagaggaagaaaccccaccgccg 640
           |||||| ||||||  | ||| || ||||| ||  | || ||| | ||    |    ||
Db   76916 TCTTTTACAAACGTGGCTTACGAAAGGTTGCCTTTGGACGAGCATGAGGAGCACTTGCAA 76975

Qy     641 agaaccaccgggg 653
           ||   |  ||| |
Db   76976 AGTGGCGGCGGCG 76988


RESULT   6
AC002387
LOCUS       AC002387               122871 bp    DNA     linear   PLN 05-APR-2000
DEFINITION  Arabidopsis thaliana chromosome II section 242 of 255 of the
            complete sequence. Sequence from clones T14P1, F4L23.
ACCESSION   AC002387 AE002093
VERSION     AC002387.2  GI:6598365
KEYWORDS    HTG.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 122871)
  AUTHORS   Lin,X., Kaul,S., Rounsley,S.D., Shea,T.P., Benito,M.-I., Town,C.D.,
            Fujii,C.Y., Mason,T.M., Bowman,C.L., Barnstead,M.E.,
            Feldblyum,T.V., Buell,C.R., Ketchum,K.A., Lee,J.J., Ronning,C.M.,
            Koo,H., Moffat,K.S., Cronin,L.A., Shen,M., VanAken,S.E., Umayam,L.,
            Tallon,L.J., Gill,J.E., Adams,M.D., Carrera,A.J., Creasy,T.H.,
            Goodman,H.M., Somerville,C.R., Copenhaver,G.P., Preuss,D.,
            Nierman,W.C., White,O., Eisen,J.A., Salzberg,S.L., Fraser,C.M. and
            Venter,J.C.
  TITLE     Sequence and analysis of chromosome 2 of the plant Arabidopsis
            thaliana
  JOURNAL   Nature 402 (6763), 761-768 (1999)
  MEDLINE   20083487
   PUBMED   10617197
REFERENCE   2  (bases 1 to 122871)
  AUTHORS   Lin,X.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
COMMENT     On Dec 17, 1999 this sequence version replaced gi:2583106.
            The sequence and annotation of chromosome 2 were merged from those
            of the individual clones on this chromosome after removing
            overlaps. For detailed information, please see the TIGR web site
            (http://www.tigr.org/tdb/at/at.html).

            Genes were identified by a combination of three methods: Gene
            prediction programs including GRAIL
            (ftp://arthur.epm.ornl.gov/pub/xgrail), Genefinder (Phil Green,
            University of Washington), Genscan (Chris Burge,
            http://gnomic.stanford.edu/GENSCANW.html), and NetPlantGene
            (http://www.cbs.dtu.dk/services/NetGene2/), searches of the
            complete sequence against a peptide database and plant EST
            databases at TIGR, and manual curations based on those analyses.
            Annotated genes are named to indicate the level of evidence for
```

their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by two or more gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats were identified by repeatmasker (Arian Smit, http://ftp.genome.washington.edu/RM/RepeatMasker.html). Genes are numbered from the top to bottom of the chromosome.

We thank the CSHL/WashU/ABI consortium for sequencing BAC clones F6P23, F5J6, T17A5, and T13L16, the ESSA group for sequencing clone F13D4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards and Satoshi Tabata for helpful assistance. In addition, we would like to thank the TIGR Bioinformatics Department, especially Lixin Zhou, Hanif Khalak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy Peterson, Michael Holmes, and Delwood Richardson for software and database support.

This work was supported by the National Science Foundation, Department of Energy and the US Department of Agriculture.

Address all correspondence to: at@tigr.org.

FEATURES Location/Qualifiers

source 1. .122871
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="II"

misc_feature complement(<1. .2952)
/note="Sequence from clone T14P1"

repeat_region complement(519. .627)
/rpt_family="(TAAAA)n"

mRNA join(2227. .2322,2480. .2527,2650. .2705,3098. .3150, 3250. .3641)
/gene="At2g45170"

gene 2227. .3641
/gene="At2g45170"
/note="T14P1.2"

CDS join(2488. .2527,2650. .2705,3098. .3150,3250. .3469)
/gene="At2g45170"
/codon_start=1
/product="putative microtubule-associated protein"
/protein_id="AAF18602.1"
/db_xref="GI:6598366"
/translation="MNKGSIFKMDDDFEKRKAEAGRIREKYPDRIPVIVEKAEKSEVP NIDKKKYLVPSDLTVGQFVYVIRKRIKLSAEKAIFIFVDNVLPPTGELMSSVYEDKKD EDGFLYITYSGENTFGASSI"

misc_feature complement(2953. .>122871)
/note="Sequence from clone F4L23"

mRNA 4282. .4870
/gene="At2g45180"

gene 4282. .4870
/gene="At2g45180"
/note="F4L23.31"

CDS 4320. .4724
/gene="At2g45180"
/codon_start=1
/product="putative proline-rich protein"
/protein_id="AAB82643.1"
/db_xref="GI:2583134"
/translation="MASKALAVTALLITLNLLFFTFVTSTKCPPTTPKPPKTPKSPKK APAVKPTCPTDTLKLGVCADLLGLVNVVVGSPPKTPCCTLLQGLANLEAAVCLCTALK ANVLGINLNVPIDLTLLLNYCGKKVPHGFQCS"

repeat_region 5379. .5403
/rpt_family="POLY_A"

mRNA complement(join(<6393. .6470,6586. .6663,7082. .7157, 7248. .7296,7872. .8004,8099. .8260,8382. .>8495))
/gene="At2g45190"

gene complement(<6393. .>8495)
/gene="At2g45190"
/note="F4L23.30"

CDS complement(join(6393. .6470,6586. .6663,7082. .7157, 7248. .7296,7872. .8004,8099. .8260,8382. .8495))
/gene="At2g45190"
/note="unknown protein"
/codon_start=1
/protein_id="AAB82644.1"
/db_xref="GI:2583135"
/translation="MSMSSMSSPSSAVCSPDHFSPSDHLCYVQCNFCQTILAVNVPYT SLFKTVTVRCGCCTNLLSVNMRSYVLPASNQLQLQLGPHSYFNPQDILEELRDAPSNM NMMMMNQHPTMNDIPSFMDLHQQHEIPKAPPVNRPPEKRQRVPSAYNRFIKEEIQRIK AGNPDISHREAFSAAAKNWAHFPHIHFGLVPDNQPVKKTNMPQQEGEDNMVMKEGFYA PAAANVGVTPY"

repeat_region 10147. .10201
/rpt_family="(GA)n"

repeat_region 12613. .12646
/rpt_family="(CATA)n"

mRNA complement(join(15613. .15729,15821. .15950,16066. .16137, 16252. .16341,16810. .17006,17451. .17647))
/gene="At2g45200"

gene complement(15613. .17647)
/gene="At2g45200"
/note="F4L23.29"

CDS complement(join(15632. .15729,15821. .15950,16066. .16137, 16252. .16341,16810. .17006,17451. .17583))
/gene="At2g45200"
/codon_start=1
/product="putative cis-Golgi SNARE protein"
/protein_id="AAB82642.1"
/db_xref="GI:2583133"
/translation="MTESSLDLQESGWEELRREARKIEGDLDVKLSSYAKLGARFTQG GYVDTGSPTVGSGRSWKSMEMEIQSLLEKLLDINDSMSRCAASAAPTTSVTQKLARHR DILHEYTQEFRRIKGNINSLREHAELLSSVRDDISEYKASGSMSPGVQVLRERASIHG SISHIDDVIGQAQATRAVLGSQRSLFSDVQGKVKNLGDKFPVIRGLLGSIKRKRSRDT LILSAVIAACTLFLIIYWLSK"

misc_feature complement(17986. .18862)
/note="molecular marker ML"

repeat_region complement(18449. .18564)
/rpt_family="(TAAAA)n"

mRNA 19741. .>20315
/gene="At2g45210"

gene 19741. .>20315
/gene="At2g45210"
/note="F4L23.28; member of auxin-induced protein multigene family"

CDS 19827. .20315
/gene="At2g45210"
/codon_start=1
/product="putative auxin-induced protein"
/protein_id="AAB82641.1"
/db_xref="GI:2583132"
/translation="MRKIIGFRIGRRVSRWIFRKTRIQRSGYNRIHSTQQACMLMRPL AKLKSWGQRLKQSFRRRSTRRSAYIPVDHKKADPVPRGHLAIYVGQKDGDCHRVLVPI VYFNHPLFGELLREAEKEYGFCHEGGITIPCLYSDFERVKTRIASGSSSRVFPWGRHC RN"

mRNA complement(join(<22224. .22909,23488. .>24334))
/gene="At2g45220"

gene complement(<22224. .>24334)
/gene="At2g45220"
/note="F4L23.27"

CDS complement(join(22224. .22909,23488. .24334))
/gene="At2g45220"
/codon_start=1
/product="putative pectinesterase"
/protein_id="AAB82640.1"
/db_xref="GI:2583131"
/translation="MAFRAYIINFVILCILVASTVSGYNQKDVKAWCSQTPNPKPCEY FLTHNSNNEPIKSESEFLKISMKLVLDRAILAKTHAFTLGPKCRDTREKAAWEDCIKL YDLTVSKINETMDPNVKCSKLDAQTWLSTALTNLDTCRAGFLELGVTDIVLPLMSNNV SNLLCNTLAINKVPFNYTPPEKDGFPSWVKPGDRKLLQSSTPKDNAVVAKDGSGNFKT IKEAIDAASGSGRFVIYVKQGVYSENLEIRKKNVMLRGDGIGKTIITGSKSVGGGTTT FNSATVAAVGDGFIARGITFRNTAGASNEQAVALRSGSDLSVFYQCSFEAYQDTLYVH SNRQFYRDCDVYGTVDFIFGNAAAVLQNCNIFARRPRSKTNTITAQGRSDPNQNTGII

```
                     IHNSRVTAASDLRPVLGSTKTYLGRPWRQYSRTVFMKTSLDSLIDPRGWLEWDGNFAL
                     KTLFYAEFQNTGPGASTSGRVTWPGFRVLGSASEASKFTVGTFLAGGSWIPSSVPFTS
                     GL"
     repeat_region   23009. .23038
                     /rpt_family="(TA)n"
     repeat_region   28701. .28741
                     /rpt_family="POLY_A"
     repeat_region   complement(29132. .29228)
                     /rpt_family="POLY_A"

  Query Match             21.1%;  Score 205.2;  DB 8;  Length 122871;
  Best Local Similarity   61.1%;  Pred. No. 9.5e-47;
  Matches  352;  Conservative    0;  Mismatches  218;  Indels    6;  Gaps    1;

Qy      58 taacatggaacttaacagatctgaagcagacgaagcaaaggccgagaccactcccaccgg 117
           | | || || |  |||          ||| || |  | |       ||  ||  | ||||
Db 105952 TGATATCGATCCCAACGAGCACTCTTCAGCCGGAAAAGATCAAAGTACTCCTGGCTCCGG 106011

Qy     118 tggagccaccagctcagccacagcctctggctcttcctccggacgtcgtccacgtggtcg 177
           |||||  | | ||   |    ||      |    | |  |     |  | ||||||||  |
Db 106012 TGGAGAAAGCGGCGGCGGAGGAGGAGGAGATAATCACATCACGAGAAGGCCACGTGGCAG 106071

Qy     178 tcctgcaggttccaaaaacaaacccaaacctccgacgattataactagagatagtcctaa 237
           || || || || || |||||||| ||||| || |  || || ||| |||| ||  | ||
Db 106072 ACCAGCGGGATCTAAGAACAAACCAAAACCGCCAATCATCATCACTCGAGACAGCGCAAA 106131

Qy     238 cgtccttagatcacacgttcttgaagtcacctccggttcggacatatccgaggcagtctc 297
           ||  || | ||| || ||  | |||||  |   ||| |  ||| |    ||    ||| |
Db 106132 CGCTCTCAAATCTCATGTCATGGAAGTAGCAAACGGATGTGACGTCATGGAAAGTGTCAC 106191

Qy     298 cacctacgccactcgtcgcggctgcggcgtttgcattataagcggcacgggtgcggtcac 357
           |  || |||    ||||||    | ||| | ||| || | ||||| |  || || || ||
Db 106192 CGTCTTCGCTCGCCGTCGCCAACGTGGCATCTGCGTTTTGAGCGGAAACGGCGCCGTTAC 106251

Qy     358 taacgtcacgatacggcaacctgcggctccggctggtggagg------tgtgattaccct 411
           ||||| || ||| | ||||| || |     |||||| ||      |||  ||| | |
Db 106252 CAACGTTACCATAAGACAACCAGCTTCAGTACCTGGTGGTGGCTCATCTGTCGTTAACTT 106311

Qy     412 gcatggtcggtttgacattttgtctttgaccggtactgcgcttccaccgcctgcaccacc 471
           || || || || || ||| | ||| |  | ||  |    ||||| || || || |||||
Db 106312 ACACGGACGTTTCGAGATTCTTTCTCTCTCGGGATCATTCCTTCCTCCTCCGGCTCCACC 106371

Qy     472 gggagcaggaggtttgacggtgtatctagccggaggtcaaggacaagttgtaggagggaa 531
           |  ||   |||| | ||| | ||  ||||||| ||||| ||||| ||||| ||||| |
Db 106372 AGCTGCGTCAGGTCTAACGATTTACTTAGCCGGTGGTCAGGGACAGGTTGTTGGAGGAAG 106431

Qy     532 tgtggctggttcgttaattgcttcgggaccggtagtgttgatggctgcttcttttgcaaa 591
           |||| |||| |  | || ||||| ||||| |||||| | ||||| |||||  |||| |||
Db 106432 CGTGGTTGGTCCACTCATGGCTTCAGGACCTGTAGTGATTATGGCAGCTTCGTTTGGAAA 106491

Qy     592 cgcagtttatgataggttaccgattgaagaggaaga 627
           ||| |  ||||| ||  | ||| | || || || ||
Db 106492 CGCTGCGTATGAGAGACTGCCGTTGGAGGAAGACGA 106527


RESULT   7
AC015450
LOCUS       AC015450               100806 bp    DNA     linear   PLN 15-JUN-2001
DEFINITION  Arabidopsis thaliana chromosome 1 BAC F14G6 genomic sequence,
            complete sequence.
ACCESSION   AC015450
VERSION     AC015450.5  GI:12323968
KEYWORDS    HTG.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 100806)
  AUTHORS   Lin,X., Kaul,S., Town,C.D., Benito,M.-I., Creasy,T.H., Haas,B.J.,
            Wu,D., Maiti,R., Ronning,C.M., Koo,H., Fujii,C.Y., Utterback,T.R.,
            Barnstead,M.E., Bowman,C.L., White,O., Nierman,W.C. and Fraser,C.M.
  TITLE     Arabidopsis thaliana chromosome 1 BAC F14G6 genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 100806)
  AUTHORS   Lin,X. and Kaul,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1999) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org
REFERENCE   3  (bases 1 to 100806)
  AUTHORS   Town,C.D. and Kaul,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JAN-2001) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
COMMENT     On Jan 19, 2001 this sequence version replaced gi:12280761.
            Address all correspondence to:at@tigr.org

            BAC clone F14G6 is from Arabidopsis thaliana chromosome 1
            The orientation of the sequence is from SP6 to T7 end of the BAC
            clone.
            Genes were identified by a combination of several methods: Gene
            prediction programs including Genscan+ (Chris Burge,
            http://CCR-081.mit.edu/GENSCAN.html), GeneMarkHMM (Mark Borodovsky,
            http://genemark.biology.gatech.edu/GeneMark/), GlimmerA (a variant
            of GlimmerM, see Mihaela Pertea,
            http://www.tigr.org/softlab/glimmerm_htm/glimmerm.html, and
            GeneSplicer (Mihaela Pertea and Steven Salzberg, contact
            mpertea@tigr.org), searches of the complete sequence against a
            peptide database and the plant EST database at TIGR
            (http://www.tigr.org/tdb/tgi.shtml). Annotated genes are named to
            indicate the level of evidence for their annotation. Genes with
            similarity to other proteins are named after the database hits.
            Genes without significant peptide similarity but with EST
            similarity are named as  unknown  proteins. Genes without protein
            or EST similarity, that are predicted by more than two gene
            prediction programs over most of their length are annotated as
            hypothetical  proteins. Genes encoding tRNAs are predicted by
            tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
            Simple repeats are identified by repeatmasker (Arian Smit,
            http://ftp.genome.washington.edu/RM/RepeatMasker.html).
FEATURES             Location/Qualifiers
     source          1. .100806
                     /organism="Arabidopsis thaliana"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="1"
                     /clone="F14G6"
     misc_feature    1
                     /note="35000 nt  before this point were not included in
                     the submitted sequence, due to overlap with another BAC
                     [F15M4]"
     mRNA            <5217. .>5774
                     /gene="F14G6.1"
     gene            5217. .5774
                     /gene="F14G6.1"
                     /note="contains Pfam profile: PF00097 Zinc finger, C3HC4
                     type (RING finger)"
     CDS             5217. .5774
                     /gene="F14G6.1"
                     /codon_start=1
                     /product="putative RING zinc finger protein; 5217-5774"
                     /protein_id="AAG51946.1"
                     /db_xref="GI:12323975"
                     /translation="MARLLFRLLQEANSTSPAEASPPFNSDLVLILAVLLCALTCIIG
                     LIAVSRCAWLRRIASRNRSDQTHPPPVAAANKGLKKKVLRSLPKLTYSPDSPPAEKLV
                     ECAICLTEFAAGDELRVLPQCGHGFHVSCIDTWLGSHSSCPSCRQILVVTRCHKCGGL
                     PGSSSSGPEPDTRIKQREDGPDNLP"
     mRNA            complement(join(<8331. .8855,9526. .9800,9933. .>10137))
                     /gene="F14G6.2"
     gene            complement(8331. .10137)
                     /gene="F14G6.2"
                     /note="N-term similar to N-term of NAM GB:CAA63101
                     [Petunia x hybrida] (apical meristem formation), CUC2
                     GB:BAA19529 [Arabidopsis thaliana],  GRAB2 protein
                     GB:CAA09372 [Triticum sp.]"
     CDS             complement(join(8331. .8855,9526. .9800,9933. .10137))
```

/gene="F14G6.2"
/codon_start=1
/product="unknown protein; 10137-8331"
/protein_id="AAG51953.1"
/db_xref="GI:12323982"
/translation="MMLAVEDVLSELAGEERNERGLPPGFRFHPTDEELITFYLASKI
FHGGLSGIHISEVDLNRCEPWELPEMAKMGEREWYFYSLRDRKYPTGLRTNRATTAGY
WKATGKDKEVFSGGGGQLVGMKKTLVFYKGRAPRGLKTKWVMHEYRLENDHSHRHTCK
EEWVICRVFNKTGDRKNVGLIHNQISYLHNHSLSTTHHHHHEALPLLIEPSNKTLTNF
PSLLYDDPHQNYNNNNFLHGSSGHNIDELKALINPVVSQLNGIIFPSGNNNNDEDDFD
FNLGVKTEQSSNGNEIDVRDYLENPLFQEASYGLLGFSSSPGPLHMLLDSPCPLGFQL
"
mRNA complement(join(<15417. .16305,17467. .>18176))
/gene="F14G6.3"
gene complement(15417. .18176)
/gene="F14G6.3"
/note="similar to phosphate transporter GB:BAA20522
[Catharanthus roseus]; contains Pfam profile: PF00083
Sugar (and other) transporter"
CDS complement(join(15417. .16305,17467. .18176))
/gene="F14G6.3"
/codon_start=1
/product="putative phosphate transporter; 18176-15417"
/protein_id="AAG51941.1"
/db_xref="GI:12323970"
/translation="MPELSLLSALDAARIQWYHFKAIIVAGMGLFTDAYDLFCIAPIM
KMISQIYYHKDSIGTALLSTSYAIALLGTALGQLIFGYLGDRVGRRKVYGLSLLIMVF
SSFGCGFSVCTTRRSCVMVSLGFFRFVLGLGIGGDYPLSATIMSEFANKRTRGAFIAA
VFSMQGLGILMSSAVTMVVCLAFKNAGEGSSEKTNVAGLETLAPPESDIAWRLILMIG
ALPAALTFYWRMLMPETARYTALVENNVVQAAKDMQRVMSVSMISQITEDSSSELEQP
PSSSSYKLFSRRFLSLHGRDLFAASANWFLVDVVFYTSNLLLSQIFNFSNKPLNSTNV
YDSAFEVAKLAAIVAACSTIPGYWFTVYFIDKIGRVKIQMMGFFLMAVVYLVAGIPYS
WYWSKHEKTNKGFMVLYGLIFFFSNFGPNTTTFIIPAELFPARFRSTCHGISGAAGKF
GAIVGTVGFLWATRHHEEDGFPDVKRVRIAFLILGGVCIAGMIVTYLFTRETMGRSLE
ENEDEIVSTSAGSSPANELLRRQY"
mRNA complement(join(<18673. .18926,19124. .>19301))
/gene="F14G6.4"
gene complement(18673. .19301)
/gene="F14G6.4"
/note="predicted by genscan+"
CDS complement(join(18673. .18926,19124. .19301))
/gene="F14G6.4"
/codon_start=1
/product="hypothetical protein; 19301-18673"
/protein_id="AAG51943.1"
/db_xref="GI:12323972"
/translation="MNAENNQTTTHSKVISHVFCTGTAKLGSVGPPIGLVDIGVSEV
AYIFRVSLPGIEKNQDKIKCEIQREGRVCIQGVIPEIAIPSDTGCLYRMQVQQLCPPG
PFSITFNLPGQVDPRLFSPNFRSDGIFEVVVVKLGVRIPTS"
mRNA join(<20920. .20986,21073. .21148,21227. .21330,21407. .21472,
21739. .21821,21911. .21986,22064. .22182,22264. .22308,
22385. .>22411)
/gene="F14G6.5"
gene 20920. .22411
/gene="F14G6.5"
CDS join(20920. .20986,21073. .21148,21227. .21330,21407. .21472,
21739. .21821,21911. .21986,22064. .22182,22264. .22308,
22385. .22411)
/gene="F14G6.5"
/codon_start=1
/product="unknown protein; 20920-22411"
/protein_id="AAG51945.1"
/db_xref="GI:12323974"
/translation="MAAISPWLSSPQSFSNPRVTITDSRRCSSISAAISVLDSSNEEQ
HRISSRDHVGMKRRDVMLQIASSVFFLPLAISPAFAETNASEAFRVYTDETNKFEISI
PQVSIAITGLGPDFTRMESFGKVEAFAETLVSGLDRSWQKPVGVTAKLIDSRASKGFY
YIEYTLQNPGEARKHLYSAIGMATNGWYNRLYTVTGQFTDEESAEQSSKIQKTVKSFR
FI"
mRNA complement(join(<23340. .23431,23548. .23785,23879. .23919,
24008. .24124,24234. .24378,24602. .>24808))
/gene="F14G6.6"
gene complement(23340. .24808)
/gene="F14G6.6"
/note="contains Pfam profile: PF00076 RNA recognition
motif. (a.k.a. RRM, RBD, or RNP domain)"
CDS complement(join(23340. .23431,23548. .23785,23879. .23919,
24008. .24124,24234. .24378,24602. .24808))
/gene="F14G6.6"
/codon_start=1
/product="putative RNA-binding protein; 24808-23340"
/protein_id="AAG51948.1"
/db_xref="GI:12323977"
/translation="MAYQPVPGSGFHYLNSPFGDTTFTKVFVGGLAWETQSETLRQHF
EQYGEILEAVVIADKNTGRSKGYGFVTFRDPEAARRACADPTPIIDGRRANCNLASLG
RPRPPLPYAVIPNMPDLLPRTLEMCRVLEVHLLEIIRISNHFHITTSKELSILMVTAY
GPEYMYSQSQGLYSPYMGQQYLQVYGVPGAVNSPVYQYGQLSQTIPNGHGYTAVQGYS
VPGSHILQLGGPTVSTMTTSSMPALQAPYPSGIPGPAPVQSHIIVHSPQFMQSTASDQ
TTR"
mRNA complement(join(<26257. .26401,26478. .26667,26773. .26935,
27062. .27247,27332. .27486,27593. .>27736))
/gene="F14G6.7"
gene complement(26257. .27736)
/gene="F14G6.7"
/note="similar to cinnamoyl-CoA reductase GB:CAA56103
[Eucalyptus gunnii]; contains non-consensus GG acceptor
splice site at exon 4"
CDS complement(join(26257. .26401,26478. .26667,26773. .26935,
27062. .27247,27332. .27486,27593. .27707))
/gene="F14G6.7"
/note="contains nonconsensus splice site"
/codon_start=1
/product="putative cinnamoyl-CoA reductase; 27707-26257"
/protein_id="AAG51951.1"
/db_xref="GI:12323980"
/translation="MAVKQKVCVTGAGGFIASWLVKFLLSRGYTVHGTVRDPCDEKND
HLRKLDNASKNLKLFKADLFDDEGLFSAIDGCSGVFHIASPVPFEGEELIKPALTGTK
NVLEACTETKVQKVVVVSSIAAVVYNPKWPQDVAKDEDCWSDTQYLHSLERYYYLAKT
LTEREALEWSKRNFADVVTLCPSVIIGPRLQSTLNSSSLGLLKFIKGGIKSLLSDELY
LVDVRDVADALLLVYENREATGRYICNSHSLYTDSLMEKLKNMYPKRNFPESFTEVKE
KEVRPLSAEKLKNLGWKFRPLEETIDDSVVSFEAAGDLPKA"
mRNA complement(join(<28495. .28664,28744. .28834,28902. .29033,

Query Match 20.7%; Score 201.8; DB 8; Length 100806;
Best Local Similarity 63.9%; Pred. No. 8.6e-46;
Matches 333; Conservative 0; Mismatches 167; Indels 21; Gaps 1;

```
Qy    129 gctcagccacagcctctggctcttcctccggacgtcgtccacgtggtcgtcctgcaggtt 188
          | || | | |||    |       |||  ||||    ||||||||||| |||||| | || |
Db  42015 GTTCCGACCCAGTTACCTCTGGTTCAACCGGGAAACGTCCACGTGGACGTCCTCCGGGAT 42074

Qy    189 ccaaaaacaaacccaaacctccgacgattataactagagatagtcctaacgtccttagat 248
          |||| ||||| || || || |||  |||  | ||||||||||| || ||||| |||||||
Db  42075 CCAAGAACAAGCCGAAGCCACCGGTGATAGTGACTAGAGATAGCCCCAACGTGCTTAGAT 42134

Qy    249 cacacgttcttgaagtcacctccggttcggacatatccgaggcagtctccacctacgcca 308
          | || |||||||||||| | || ||  | ||||||  ||||   ||  |||| |||||
Db  42135 CTCATGTTCTTGAAGTCTCATCTGGAGCCGACATAGTCGAGAGCGTTACCACTTACGCTC 42194

Qy    309 ctcgtcgcggctgcggcgtttgcattataagcggcacgggtgcggtcactaacgtcacga 368
          |  | ||  | || || | |||| | || || |  ||  ||||  |||| ||||||||
Db  42195 GCAGGAGAGGAAGAGGAGTCTCCATTCTCAGTGGTAACGGCACGGTGGCTAACGTCAGTC 42254

Qy    369 tacggcaacctgcggctccggctggt----------------------ggaggtgtgatta 407
          | ||||| || ||  |  | || | |                     ||||||||  |
Db  42255 TCCGGCAGCCGGCAACGACAGCGGCTCATGGGGCAAATGGTGGAACCGGAGGTGTTGTGG 42314

Qy    408 ccctgcatggtcggtttgacattttgtctttgaccggtactgcgcttccaccgcctgcac 467
          | || |||||  |||||||| ||  | ||  | || ||||| | | | || || ||||| |
Db  42315 CTCTACATGGAAGGTTTGAGATACTTTCCCTCACAGGTACGGTGTTGCCGCCCCCTGCGC 42374

Qy    468 caccgggagcaggaggtttgacggtgtatctagccggaggtcaaggacaagttgtaggag 527
          | || ||| | || ||| |  |  | | |||  |||| | |||||| || ||  | ||||
Db  42375 CGCCAGGATCCGGTGGTCTTTCTATCTTTCTTTCCGGCGTTCAAGGTCAGGTGATTGGAG 42434

Qy    528 ggaatgtggctggttcgttaattgcttcgggaccggtagtgttgatggctgcttcttttg 587
          | || ||||  | | || |  | |||||||| || ||  |  | |||||||| || ||
```

```
Db  42435 GAAACGTGGTGGCTCCGCTTGTGGCTTCGGGTCCAGTGATACTAATGGCTGCATCGTTCT 42494

Qy    588 caaacgcagtttatgataggttaccgattgaagaggaagaa 628
          | || |||  ||  || ||| | ||  |||||||  |||| |
Db  42495 CTAATGCAACTTTCGAAAGGCTTCCCCTTGAAGATGAAGGA 42535


RESULT   8
ATF16J13/c
LOCUS       ATF16J13              107600 bp    DNA     linear   PLN 14-APR-1999
DEFINITION  Arabidopsis thaliana DNA chromosome 4, BAC clone  F16J13 (ESSA
            project).
ACCESSION   AL049638
VERSION     AL049638.1  GI:4586098
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 107600)
  AUTHORS   Bevan,M., Hilbert,H., Braun,M., Holzer,E., Brandt,A.,
            Duesterhoeft,A., Bancroft,I., Mewes,H.W., Mayer,K.F.X. and
            Schueller,C.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 107600)
  AUTHORS   EU Arabidopsis sequencing,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-APR-1999) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            schuelle@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
COMMENT     Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
            and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.
FEATURES             Location/Qualifiers
     source          1. .107600
                     /organism="Arabidopsis thaliana"
                     /variety="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="4"
     gene            1491. .2285
                     /gene="F16J13.10"
     exon            1491. .2285
                     /gene="F16J13.10"
                     /number=1
     CDS             1491. .2285
                     /gene="F16J13.10"
                     /note="similarity to predicted proteins Arabidopsis
                     thaliana"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAB40935.1"
                     /db_xref="GI:4586099"
                     /translation="MFSGEGASQIRIAPEDKPIQWCYQILKSRDFKSARYITQMNLKL
                     SKTRHDEYEKGLAICDILIAAENRLPNGLLDCYGMIRMTRPGLVLYQNIEKELNLLGW
                     GNISNPFPFRQEASEKFFFAWSLLSNPTIKEMYDYATSDEVNLEPQGNVNEYMDVDSS
                     SQSGYGLGSVLCSDLPGSEYLKEFADVYPLPLAEKGQAPHNRFGWYDHNVGPETNNNV
                     VVTYEDAKEEECDMSSSSELRIINGRRVKITIEEAAETSDTPSCSR"
     repeat_region   6483. .6506
                     /note="24bp A tandem repeat"
     gene            complement(6594. .7577)
                     /gene="F16J13.20"
     gene            6594. .7577
                     /gene="F16J13.20"
     CDS             complement(6594. .7577)
                     /gene="F16J13.20"
                     /note="strong similarity to other hypothetical proteins -
                     Arabidopsis thaliana"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAB40936.1"
                     /db_xref="GI:4586100"
                     /translation="MGSIPFIFIILLFHVHKSKSQTIESAHFLDLMIRDYTIRNFNIH
                     FKTGAIQKVHLPSNFSSIDIATAKFRCGSLRRHGARIGEFHLGPGLTVEPCVERVILV
                     RQNLGFNWSSYIYSTGYNLTGYKYRLVSPVLGLLAYNSNPDGVAVNPYEVNVMGTEQN
                     PILIKFLSSEASGSPKPNTKKNSSVLCACFTSNGNITFREQVSAYVCLGTRQGHYALV
                     IRAHDSGGGGSTVVTPSSSPALTDGGGGKLSRWKVAVGSVIGSIIGAFLLGLLVVAMV
                     VKGKKKAMREEMERRAYEEEALQVSMVGHVRANPNASRSRTIPRFENTRYK"
     exon            complement(6594. .7577)
                     /gene="F16J13.20"
                     /number=1
     gene            8658. .11027
                     /gene="F16J13.30"
     exon            complement(8658. .8727)
                     /gene="F16J13.30"
                     /number=1
     gene            complement(join(8658. .8727,8813. .8910,9144. .9254,
                     9361. .9412,9514. .9665,9717. .9773,9878. .9978,10067. .10163,
                     10251. .10363,10862. .11027))
                     /gene="F16J13.30"
     CDS             complement(join(8658. .8727,8813. .8910,9144. .9254,
                     9361. .9412,9514. .9665,9717. .9773,9878. .9978,10067. .10163,
                     10251. .10363,10862. .11027))
                     /gene="F16J13.30"
                     /note="strong similarity to hypothetical protein F7H19.70
                     - Arabidopsis thaliana, PID:e1310057"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAB40937.1"
                     /db_xref="GI:4586101"
                     /translation="MAFTLTIPRFSAISRKPITCSSSRTQCPAPFTHGRSISLRRRLT
                     LLPLKASTDQSGQVGGEEVDSKILPYCSINKNEKRTIGEMEQEFLQAMQSFYYEGKAI
                     MSNEEFDNLKEELMWEGSSVVMLSSDEQRFLEASMAYVSGNPILSDEEYDKLKMKLKM
                     DGSEIVCEGPRCSLRSKKNITLDSLFGTNSRDVIDLWTSLIMQVYSDLAIDYFKMFLL
                     NVPATVVALGLFFFLDDITGFEITYLLELPEPFSFIFTWFAAVPAIVYLALSLTKLIL
                     KDFLILKGPCPNCGTENVSFFGTILSIPNDSNTNNVKCSGCGTEMVYDSGSRLITLPE
                     GGKA"
     intron          complement(8728. .8812)
                     /gene="F16J13.30"
                     /number=1
     exon            complement(8813. .8910)
                     /gene="F16J13.30"
                     /number=2
     intron          complement(8911. .9143)
                     /gene="F16J13.30"
                     /number=2
     exon            complement(9144. .9254)
                     /gene="F16J13.30"
                     /number=3
     intron          complement(9255. .9360)
                     /gene="F16J13.30"
                     /number=3
     exon            complement(9361. .9412)
                     /gene="F16J13.30"
                     /number=4
     intron          complement(9413. .9513)
                     /gene="F16J13.30"
                     /number=4
     exon            complement(9514. .9665)
                     /gene="F16J13.30"
                     /number=5
     intron          complement(9666. .9716)
                     /gene="F16J13.30"
                     /number=5
     exon            complement(9717. .9773)
                     /gene="F16J13.30"
                     /number=6
     intron          complement(9774. .9877)
                     /gene="F16J13.30"
                     /number=6
     exon            complement(9878. .9978)
                     /gene="F16J13.30"
                     /number=7
```

```
     intron          complement(9979. .10066)
                     /gene="F16J13.30"
                     /number=7
     exon            complement(10067. .10163)
                     /gene="F16J13.30"
                     /number=8
     intron          complement(10164. .10250)
                     /gene="F16J13.30"
                     /number=8
     exon            complement(10251. .10363)
                     /gene="F16J13.30"
                     /number=9
     intron          complement(10364. .10861)
                     /gene="F16J13.30"
                     /number=9
     exon            complement(10862. .11027)
                     /gene="F16J13.30"
                     /number=10
     repeat_region   11789. .11824
                     /note="36bp AT tandem repeat"
     gene            14541. .16583
                     /gene="F16J13.40"
     CDS             join(14541. .14631,14724. .14959,15052. .15303,15445. .15521,
                     15593. .15914,16264. .16324,16543. .16583)
                     /gene="F16J13.40"
                     /note="EST GB:Aa651292 covers 3'end, one intron in 3'
                     untranslated region
                     similarity to RNA splicing-related protein - Rattus
                     norvegicus, PID:d1024790
                     contains EST gb:H37320, Aa651292"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAB40938.1"
                     /db_xref="GI:4586102"
                     /translation="MSSDTAKENASVVDSSLTDWKQDLGNSDDPESTSYRSKEDHKLS
                     KVDVDRRNFPDQLESAKANKNSKPGYRTRYFIIKSLNYDNIQVSVEKGIWATQVMNEP
                     ILEGAFHKSGRVILIFSVNMSGFFQGYAEMLSPVGWRRDQIWSQGGGKNNPWGRSFKV
                     KWLRLSELPFQKTLHLKNPLNDYKPVKISRDCQELPEDIGEALCELLDANSCDDGLLN
                     SSSRDDYSTKRSRAEPPSSSGDDEYNNNLWGHTPMSYPPTVYPNQDDLSRFHLAQQRG
                     YGVSPEYLHTSPGASNSRTEQDKSLRFNSWCLPLESPLANSLTDDDFLEMSYEEYVET
                     HGKCMKQLGLPVSDSTITINTRAIK"
     exon            14541. .14631
                     /gene="F16J13.40"
                     /number=1
     intron          14632. .14723
                     /gene="F16J13.40"
                     /number=1
     exon            14724. .14959
                     /gene="F16J13.40"
                     /number=2
     intron          14960. .15051
                     /gene="F16J13.40"
                     /number=2
     exon            15052. .15303
                     /gene="F16J13.40"
                     /number=3
     intron          15304. .15444
                     /gene="F16J13.40"
                     /number=3
     exon            15445. .15521
                     /gene="F16J13.40"
                     /number=4
     intron          15522. .15592

  Query Match             20.2%;  Score 196.8;  DB 8;  Length 107600;
  Best Local Similarity   64.6%;  Pred. No. 2.2e-44;
  Matches  310;  Conservative    0;  Mismatches  167;  Indels    3;  Gaps    1;

Qy    160 acgtcgtccacgtggtcgtcctgcaggttccaaaaacaaacccaaacctccgacgattat 219
          | |  | ||| | ||  | || ||||| ||||| |||||||| ||| |||| |  || ||
Db  54126 AAGAAGGCCAAGAGGAAGACCAGCAGGATCCAAGAACAAACCTAAAGCTCCAATAATCAT 54067

Qy    220 aagtagagatagtcctaacgtccttagatcacacgttcttgaagtcacctccggttcgga 279
          ||| ||||| ||  | ||||  ||| || | |||||  | ||  |      ||| |  ||
Db  54066 AACAAGAGACAGCGCAAACGCGCTTCGAACTCACGTCATGGAGATAGGAGACGGATGTGA 54007

Qy    280 catatccgaggcagtctccacctacgccactcgtcgcggctgcggcgtttgcattataag 339
          ||||   ||      |  | || | ||| |  || |||    | |||||||| |||| ||
Db  54006 CATAGTTGACTGTATGGCTACGTTCGCTAGACGCCGCCAAAGAGGCGTTTGCGTTATGAG 53947

Qy    340 cggcacgggtgcggtcactaacgtcacgatacggcaacctgcggctccggctggtggagg 399
          ||| || ||    || |||||||||| ||||| || ||||   | ||  |   |||
Db  53946 CGGTACAGGAAGCGTTACTAACGTCACTATACGTCAGCCTGGATCGCCACC---TGGCTC 53890

Qy    400 tgtgattaccctgcatggtcggtttgacattttgtctttgaccggtactgcgcttccacc 459
          ||| ||| ||| || || |||||||| ||  | ||| |  | ||  ||    | || ||
Db  53889 GGTGGTTAGCCTTCACGGCCGGTTTGAAATCCTCTCTCTTTCGGGATCTTTCTTGCCTCC 53830

Qy    460 gcctgcaccaccgggagcaggaggtttgacggtgtatctagccggaggtcaaggacaagt 519
          |||||| || || | |||    ||  | |  || || |||||||||| ||||| || ||
Db  53829 GCCTGCGCCGCCTGCAGCCACCGGACTAAGCGTTTACCTAGCCGGAGGACAAGGGCAGGT 53770

Qy    520 tgtaggagggaatgtggctggttcgttaattgcttcgggaccggtagtgttgatggctgc 579
          || ||||| | |||||  ||  | ||  |   |||||| || || ||| | ||||| ||
Db  53769 CGTTGGAGGTAGTGTGGTGGGACCTTTGTTGTGTTCGGGTCCTGTGGTGGTTATGGCGGC 53710

Qy    580 ttcttttgcaaacgcagtttatgataggttaccgattgaagaggaagaaaccccaccgcc 639
          |||||||   || || |  || ||| | ||  | ||||| || || |  |   ||||
Db  53709 TTCTTTTAGCAATGCGGCGTACGAAAGGCTGCCTTTGGAAGAAGATGAGATGCAGACGCC 53650


RESULT   9
ATCHRIV33/c
LOCUS       ATCHRIV33             190026 bp    DNA     linear   PLN 16-MAR-2000
DEFINITION  Arabidopsis thaliana DNA chromosome 4, contig fragment No. 33.
ACCESSION   AL161533
VERSION     AL161533.2  GI:7267889
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 2832 to 3389; 3666 to 4213)
  AUTHORS   Volckaert,G., Grymonprez,B., Voet,M., Robben,J., Mewes,H.W.,
            Lemcke,K. and Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   2  (bases 123056 to 190026)
  AUTHORS   Peters,S.A., van Staveren,M., Dirkse,W., Stiekema,W., Mewes,H.W.,
            Lemcke,K. and Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 34336; 23831 to 131430)
  AUTHORS   Hilbert,H., Braun,M., Holzer,E., Brandt,A., Duesterhoeft,A.,
            Mewes,H.W., Lemcke,K. and Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   4  (bases 1 to 190026)
  AUTHORS   EU Arabidopsis sequencing,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
COMMENT     Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
            and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
            this fragment has an overlap with ATCHRIV32 at the 5' end and an
            overlap with ATCHRIV34 at the 3' end.
FEATURES             Location/Qualifiers
     source          1. .190026
                     /organism="Arabidopsis thaliana"
                     /variety="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="4"
     gene            5417. .6920
```

```
                /gene="AT4g11890"
exon            5417. .5486
                /gene="AT4g11890"
                /number=1
CDS             join(5417. .5486,5584. .5675,5763. .5973,6065. .6302,
                6396. .6534,6612. .6920)
                /gene="AT4g11890"
                /note="strong similarity to KI domain interacting kinase 1
                -Zea mays, PIR2:T02053
                Contains Protein kinases signatures and profile
                AA44-66;Protein kinases signatures and profile AA159-171"
                /codon_start=1
                /product="protein kinase-like protein"
                /protein_id="CAB78232.1"
                /db_xref="GI:7267890"
                /translation="MAVTSLLDTVFRRRKKKSTEFISSVFEFDLDTIKAATNDFSELV
                GRGGFGFVYKGRLQNGQEIAVKILSTSSIRTERQFHNELIILSKLKHKNLINLLGFCT
                KRDQHGLVYEFMPNSSLDCFILDPHRAAQLNWEMCRNIIDGIARGLRYLHEESGLWVV
                HRDIKPGNILLDSDLKPKIVGFELARTMQQGENAAETTEIVGTVGYLDPEYIRSGRVS
                VKSDVYAFGVTILTIISRRKAWSVDGDSLIKYVRRCWNRGEAIDVIHEVMREEEREYS
                ISEILRYIHIALLCVDENAERRPNIDKVLHWFSCFSTPLPDPTFGNRFLVEEETNWPW
                SPSLSPGHSSVTSPISSR"
intron          5487. .5583
                /gene="AT4g11890"
                /number=1
exon            5584. .5675
                /gene="AT4g11890"
                /number=2
intron          5676. .5762
                /gene="AT4g11890"
                /number=2
exon            5763. .5973
                /gene="AT4g11890"
                /number=3
intron          5974. .6064
                /gene="AT4g11890"
                /number=3
exon            6065. .6302
                /gene="AT4g11890"
                /number=4
intron          6303. .6395
                /gene="AT4g11890"
                /number=4
exon            6396. .6534
                /gene="AT4g11890"
                /number=5
intron          6535. .6611
                /gene="AT4g11890"
                /number=5
exon            6612. .6920
                /gene="AT4g11890"
                /number=6
gene            7389. .10690
                /gene="AT4g11900"
exon            complement(7389. .7679)
                /gene="AT4g11900"
                /number=1
gene            complement(join(7389. .7679,7778. .7928,8208. .8445,
                8555. .8765,8889. .9043,9121. .9255,9322. .10690))
                /gene="AT4g11900"
CDS             complement(join(7389. .7679,7778. .7928,8208. .8445,
                8555. .8765,8889. .9043,9121. .9255,9322. .10690))
                /gene="AT4g11900"
                /note="strong similarity to KI domain interacting kinase 1
                - Zea mays, PID:g2735017
                Contains Prokaryotic membrane lipoprotein lipid attachment
                site AA472-482;Protein kinases signatures and profile
                AA658-670
                contains EST gb:AI992820.1, T20644, AA586217"
                /codon_start=1
                /product="KI domain interacting kinase 1-like protein"
                /protein_id="CAB78233.1"
                /db_xref="GI:7267891"
                /translation="MQICKKNVFLLYYGVLVFLSFQVSSSTDTISTNQPLSGFETIVS
                SGDIFELGLFTPTPDTYDHRNYYIGMWYRHVSPQTIVWVANRESPLGGDASTYLLKIL
                DGNLILHDNISATRKSHTEGTSRRSPQKISEGNLLFHETVWSTGVNSSMSKDVQAVLF
                DSGNLVLRDGPNSSAAVLWQSFDHPSDTWLPGGKIRLGSQLFTSWESLIDPSPGRYSL
                EFDPKLHSLVTVWNRSKSYWSSGPLYDWLQSFKGFPELQGTKLSFTLNMDESYITFSV
                DPQSRYRLVMGVSGQFMLQVWHVDLQSWRVILSQPDNRCDVYNSCGSFGICNENREPP
                PCRCVPGFKREFSQGSDDSNDYSGGCKRETYLHCYKRNDEFLPIENMKLATDPTTASV
                LTSGTFRTCASRCVADCSCQAYANDGNKCLVWTKDAFNLQQLDANKGHTFFLRLASSN
                ISTANNRKTEHSKGKSIVLPLVLASLVATAACFVGLYCCISSRIRRKKKQRDEKHSRE
                LLEGGLIDDAGENMCYLNLHDIMVATNSFSRKKKLGEGGFGPVYKGKLPNGMEVAIKR
                LSKKSSQGLTEFKNEVVLIIKLQHKNLVRLLGYCVEGDEKLLIYEYMSNKSLDGLLFD
                SLKSRELDWETRMKIVNGTTRGLQYLHEYSRLRIIHRDLKASNILLDDEMNPKISDFG
                TARIFGCKQIDDSTQRIVGTFGYMSPEYALGGVISEKSDIYSFGVLLLEIISGKKATR
                FVHNDQKHSLIAYEWESWCETKGVSIIDEPMCCSYSLEEAMRCIHIALLCVQDHPKDR
                PMISQIVYMLSNDNTLPIPKQPTFSNVLNGDQQLDYVFSINEATQTELEAR"
intron          complement(7680. .7777)
                /gene="AT4g11900"
                /number=1
exon            complement(7778. .7928)
                /gene="AT4g11900"
                /number=2
intron          complement(7929. .8207)
                /gene="AT4g11900"
                /number=2
exon            complement(8208. .8445)
                /gene="AT4g11900"
                /number=3
intron          complement(8446. .8554)
                /gene="AT4g11900"
                /number=3
exon            complement(8555. .8765)
                /gene="AT4g11900"
                /number=4
intron          complement(8766. .8888)
                /gene="AT4g11900"
                /number=4
exon            complement(8889. .9043)
                /gene="AT4g11900"
                /number=5
intron          complement(9044. .9120)
                /gene="AT4g11900"
                /number=5
exon            complement(9121. .9255)
                /gene="AT4g11900"
                /number=6
intron          complement(9256. .9321)
                /gene="AT4g11900"
                /number=6
exon            complement(9322. .10690)
                /gene="AT4g11900"
                /number=7
gene            13583. .16588
                /gene="AT4g11910"
exon            13583. .13711
                /gene="AT4g11910"
                /number=1
CDS             join(13583. .13711,13978. .14151,14239. .14406,14643. .14879,
                15546. .15719,15829. .15996,16238. .16588)
                /gene="AT4g11910"
                /note="similarity to hypothetical protein F7H19.100
                -Arabidopsis thaliana, PID:e1310060
                Contains Zinc finger, C2H2 type, domain AA420-440
                contains EST gb:Z34017, Z34018"
                /codon_start=1
                /product="putative protein"
                /protein_id="CAB78234.1"
                /db_xref="GI:7267892"
                /translation="MCSLATNLLLPSKMKPVFPEKLSTSSLCVTTRRSKMKNRSIVPV
                ARLFGPAIFEASKLKVLFLGVDEKKHPAKLPRTYTLTHSDITAKLTLAISQSINNSQL
                QGWANKLFRDEVVGEWKKVKGKMSLHVHCHISGGHFFLNLIAKLRYYIFCKELPVVLE
                AFAHGDEYLLNNHPELQESPVWVYFHSNIPEYNKVECWGPLWEAMSQHQHDGRTHKKS
                ETLPELPCPDECKCCFPTVARLFGQAIFEASKLNVKFLGVDEKKHPPNLPRTYTFTHS
                DITAKLTLAISHSINNSQLQGWANRLYRDEVVAEWRKVKSNMSLHVHCHISGDHFLLD
```

```
                    LIAELRYFIFCKELPMVLKAFVHGDENMLNNYPELHEAFVWVYFHSNIPKFNKVECWG
                    RLCEATSHDGCKTPTCEILPEPPCFDKCSCCFPTVSTIPWSHSHGCSHGEEDNNVAIA
                    GGNLLTKYMYKRSKRK"
     intron         13712. .13977
                    /gene="AT4g11910"
                    /number=1
     exon           13978. .14151
                    /gene="AT4g11910"
                    /number=2
     intron         14152. .14238
                    /gene="AT4g11910"

  Query Match             20.2%;  Score 196.8;  DB 8;  Length 190026;
  Best Local Similarity   64.6%;  Pred. No. 2.4e-44;
  Matches  310;  Conservative    0;  Mismatches  167;  Indels    3;  Gaps    1;

Qy     160 acgtcgtccacgtggtcgtcctgcaggttccaaaaacaaacccaaacctccgacgattat 219
           | |  | ||| | ||  | || ||||| ||||| |||||||| ||| |||| |  || ||
Db   77956 AAGAAGGCCAAGAGGAAGACCAGCAGGATCCAAGAACAAACCTAAAGCTCCAATAATCAT 77897

Qy     220 aactagagatagtcctaacgtccttagatcacacgttcttgaagtcacctccggttcgga 279
           ||| ||||| ||  | |||| ||| || | ||||  | ||  |       ||| |  ||
Db   77896 AACAAGAGACAGCGCAAACGCGCTTCGAACTCACGTCATGGAGATAGGAGACGGATGTGA 77837

Qy     280 catatccgaggcagtctccacctacgccactcgtcgcggctgcggcgtttgcattataag 339
           ||||   ||      |  | || | ||| |  || |||    | |||||||| |||| ||
Db   77836 CATAGTTGACTGTATGGCTACGTTCGCTAGACGCCGCCAAAGAGGCGTTTGCGTTATGAG 77777

Qy     340 cggcacgggtgcggtcactaacgtcacgatacggcaacctgcggctccggctggtggagg 399
           ||| || ||    || |||||||||| ||||| || ||||   | ||  |   |||
Db   77776 CGGTACAGGAAGCGTTACTAACGTCACTATACGTCAGCCTGGATCGCCACC---TGGCTC 77720

Qy     400 tgtgattaccctgcatggtcggtttgacattttgtctttgaccggtactgcgcttccacc 459
           ||| ||| ||| || || |||||||| ||  | ||| |  | ||  ||    | || ||
Db   77719 GGTGGTTAGCCTTCACGGCCGGTTTGAAATCCTCTCTCTTTCGGGATCTTTCTTGCCTCC 77660

Qy     460 gcctgcaccaccgggagcaggaggtttgacggtgtatctagccggaggtcaaggacaagt 519
           |||||| || || | |||    ||  | |  || || |||||||||| ||||| || ||
Db   77659 GCCTGCGCCGCCTGCAGCCACCGGACTAAGCGTTTACCTAGCCGGAGGACAAGGGCAGGT 77600

Qy     520 tgtaggagggaatgtggctggttcgttaattgcttcgggaccggtagtgttgatggctgc 579
            || ||||| | |||||  ||  | ||  |   |||||| || || ||| | ||||| ||
Db   77599 CGTTGGAGGTAGTGTGGTGGGACCTTTGTTGTGTTCGGGTCCTGTGGTGGTTATGGCGGC 77540

Qy     580 ttcttttgcaaacgcagtttatgataggttaccgattgaagaggaagaaaccccaccgcc 639
           |||||||   || || |  || || ||| | ||  | ||||| || || |  |   ||||
Db   77539 TTCTTTTAGCAATGCGGCGTACGAAAGGCTGCCTTTGGAAGAAGATGAGATGCAGACGCC 77480


RESULT  10
ATCHRIV39
LOCUS       ATCHRIV39              197976 bp    DNA     linear   PLN 16-MAR-2000
DEFINITION  Arabidopsis thaliana DNA chromosome 4, contig fragment No. 39.
ACCESSION   AL161539
VERSION     AL161539.2  GI:7268147
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 197976)
  AUTHORS   EU Arabidopsis sequencing,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
COMMENT     Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
            and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
            this fragment has an overlap with ATCHRIV38 at the 5' end and an
            overlap with ATCHRIV40 at the 3' end.
FEATURES            Location/Qualifiers
     source         1. .197976
                    /organism="Arabidopsis thaliana"
                    /variety="Columbia"
                    /db_xref="taxon:3702"
                    /chromosome="4"
     exon           complement(5668. .5772)
                    /gene="AT4g14420"
                    /number=1
     gene           5668. .7235
                    /gene="AT4g14420"
     CDS            complement(join(5668. .5772,5865. .5963,6051. .6140,
                    6238. .6352,7168. .7235))
                    /gene="AT4g14420"
                    /note="strong similarity to Nicotiana tabacum ORF able to
                    induce HR-likelesions, PATX:G1762945
                    contains EST gb:AI999161.1"
                    /codon_start=1
                    /product="elicitor like protein"
                    /protein_id="CAB78484.1"
                    /db_xref="GI:7268148"
                    /translation="MELASFLGRALFVSVFLLSAWQEFNDFGEDGGRSAKSLKPKFNA
                    FVNHVTTHTGQQLPPVDMKILVAAAIALKGIGGLLFVFGSSLGAYLLLLHQAVATPIL
                    YDFYNYDVDRKEFGQLFSKFTQSLALLGGLLFFIGMKNSRKHGRQLRKKAPKAKAN"
     gene           complement(join(5668. .5772,5865. .5963,6051. .6140,
                    6238. .6352,7168. .7235))
                    /gene="AT4g14420"
     intron         complement(5773. .5864)
                    /gene="AT4g14420"
                    /number=1
     exon           complement(5865. .5963)
                    /gene="AT4g14420"
                    /number=2
     intron         complement(5964. .6050)
                    /gene="AT4g14420"
                    /number=2
     exon           complement(6051. .6140)
                    /gene="AT4g14420"
                    /number=3
     intron         complement(6141. .6237)
                    /gene="AT4g14420"
                    /number=3
     exon           complement(6238. .6352)
                    /gene="AT4g14420"
                    /number=4
     intron         complement(6353. .7167)
                    /gene="AT4g14420"
                    /number=4
     exon           complement(7168. .7235)
                    /gene="AT4g14420"
                    /number=5
     exon           complement(8407. .9129)
                    /gene="AT4g14430"
                    /number=1
     gene           8407. .9129
                    /gene="AT4g14430"
     gene           complement(8407. .9129)
                    /gene="AT4g14430"
     CDS            complement(8407. .9129)
                    /gene="AT4g14430"
                    /note="similarity to carnitine racemase, Escherichia coli
                    contains EST gb:H36845, AI992737.1, T76583"
                    /codon_start=1
                    /product="carnitine racemase like protein"
                    /protein_id="CAB78485.1"
                    /db_xref="GI:7268149"
                    /translation="MCTLEKRGDLFLLTLTGDGEHRFHPDTIATILSLLEQAKSQSTR
                    GSILITTANGKFFSNGFDLAWAQTAGSKTGAANRLHQMVESFKPVVAALLDLPMPTIA
                    ALNGHAAAAGLILALSHDYVFMRKDRGVLYMSEVDIGLSMPDYFSALVRAKIGTSAAR
                    RELLLSGKKIRGEEAVGLGIVDSAAYDSEEGVVVASVRLGEKLAAKKWSGEVYASIRK
                    SLYPELCGILGLETRVFATPKL"
```

```
CDS             complement(10443. .11159)
                /gene="AT4g14440"
                /note="similarity to carnitine racemase, Escherichia coli
                contains EST gb:AI994308.1, AA394352, H36791, AA395864,
                T76583, H77091, T46265, AA597888, Z17630"
                /codon_start=1
                /product="carnitine racemase like protein"
                /protein_id="CAB78486.1"
                /db_xref="GI:7268150"
                /translation="MCTLEKRGDLFLLTLTGEDEHRFHPDTIASVLSLLEQAKSQSTK
                GSVLITTGHGKFFSNGFDLAWAQSAGHGAIKRMHQMVKSFKPVLAALLDLPMPTIAAL
                NGHAAASGLMFALSHDYVFMRKDRGVLYMSEVDIGLPVPDYFSALVVAKVGSGIARRE
                LLLSGKKLKGEEAVALGIVDSAAHDSAEGVVEATVSLGESLAAKKWNGEVYATIRKSL
                YPELCRMVDLTANNLATHNL"
exon            complement(10443. .11159)
                /gene="AT4g14440"
                /number=1
gene            10443. .11159
                /gene="AT4g14440"
gene            complement(10443. .11159)
                /gene="AT4g14440"
exon            13070. .13365
                /gene="AT4g14450"
                /number=1
gene            13070. .14473
                /gene="AT4g14450"
CDS             join(13070. .13365,14296. .14473)
                /gene="AT4g14450"
                /note="similarity to adenylate cyclase (EC 4.6.1.1) II,
                rat
                contains EST gb:AF067468, AI994518.1, T44124"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="CAB78487.1"
                /db_xref="GI:7268151"
                /translation="MASTMQRTSSSAASNERQLSQLQRRAPSLMIKPTSFSNWNVAIP
                LLSPLAPSLTSSFDQSHVPPPQNKTEIPVEEEVKKTPVFKKWQHPASPFCYEPTTENR
                ASRTSLFDGLDGLEEGRLRASSSYAHDERDNDEALENLQDRVSFLKRVCFFHLYN"
intron          13366. .14295
                /gene="AT4g14450"
                /number=1
exon            14296. .14473
                /gene="AT4g14450"
                /number=2
LTR             17117. .17336
misc_feature    17117. .22563
                /note="LTR retrotransposon"
gene            17575. .22327
                /gene="AT4g14460"
CDS             join(17575. .19093,19145. .19476,19552. .21256,21338. .21472,
                21549. .22327)
                /gene="AT4g14460"
                /note="similarity to simlarity to retrovirus-related
                polyprotein (retrotransposon Tal-3), Arabidopsis thaliana"
                /codon_start=1
                /product="retrovirus-related like polyprotein"
                /protein_id="CAB78488.1"
                /db_xref="GI:7268152"
                /translation="MGSYVPPPANPVTRPPINNGDQIRYPVDQYENPYYLHSADHAGL
                ILVSDRLTTASDFHSWRRSILMALNVRNKLGFINGTITKPPEDHRDFGAWSRCNDIVS
                TWLMNSVDKKIGQSLLYIATVQGIWNNLLSRFKQDDAPRIFDIEQKLSKIEQGSMDIS
                TYYTALLTLWEEHRNYVELPVCTCGRCECDAAVKWEHLQQRSRVTKFLKELNEGFDQT
                RRHILMLKPIPTIKEAFNMVTQDERQRNVKPLTRVDSVAFQNTSMINEDENAYVAAYN
                TVRPNQKPICTHCGKVGHTIQKCYKVHGYPPGMKTGNTGYTYKPNPQLHVQPRMPMMP
                QPRMQFPAQPYTNSMQKANVVAQVYAETGAYPSEGYSQAPMMNPYGSYPMPHITHGGN
                NLSLQDFTPQQIEQMISQFQAQVQVPEPAASSSNPSPLATVSEHGFMALTSTSGTIIP
                FPSTSLKYENNDLKFQNHTLSALQKFLPSDAWIIDSGASSHVCSDLAMFRELKSVSGT
                VHITQKLILHNVLHVPDFKFNLMSVSSLVKTISCSAHFYVDCCLIQELSQGLMIGRGR
                LYHNLYILETENTSPSTSTPAACLFTGSVLNDGHLWHQRLGHPSSVVLQKLKRLAYIS
                HNNLASNPFDLVHLDIWGPFSIESIEGFRYFLTVVDDCTRTTWVYMLRNKKDVSSVFP
                EFIKLVSTQFNAKIKAIRSDNAPELGFTEIVKEHGMLHHFSCAYTPQQNSVVERKHQH
                ILNVARALLFQSNIPMQYWSDCVTTAVFLINRLPSPLLNNKSPYELILNKQPDYSLLK
                NFGCLCFVSTNAHERTKFTPRARACVFLGYPSGYKGYKVLDLESHSVTVSRNVVFKEH
                VFPFKTSELLNKAVDMFPNSILPLPAPLHFVETMPLIDEDSLIPTTTDSRTADNHASS
                SSSALPSIIPPSSNTETQDIDSNAVPITRSKRTTRAPSYLSEYHCSLVPSISTLPPTD
                SSIPIHPLPEIFTASSPKKTTPYPISTVVSYDKYTPLCQSYIFAYNTETEPKTFSQAM
                KSEKWIRVAVEELQAMELNKTWSVESLPPDKNVVGCKWVFTIKYNPDGTVERYKARLV
                AQGFTQQEGIDFLDTFSPVAKLTSAKMMLGLAAITGWTLTQMDVSDAFLHGDLDEEIF
                MSLPQGYTPPAGTILPPNPVCRLLKSIYGLKQASRQWYKRFVAALVYIDDIMIASNND
                AEVENLKALLRSEFKIKDLGPARFFLGLLGCKPSSIPMDPTLHLVRDMGTPLPNPTAY
                RKLIGRLLYLTITRPDITYAVHQLSQFISAPSDIHLQAAHKVLRYIKANPGQGLMYSA
                DYEICLNGFSDADWAACKDTRRSISGFCIYLGTSLISWKSKKQAVASRSSTESEYRSM
                AQATCEIIWLQQLLKDLHIPLTCPAKLFCDNKSALHSSLNPVFHERTKHIEIDCHTVR
                DQIKAGNLKALHVPTENQHADILTKALHPGPFHHLLRQMSLSSLFLPNKPVRS"
exon            17575. .19093
                /gene="AT4g14460"
                /number=1
intron          19094. .19144
                /gene="AT4g14460"
                /number=1
exon            19145. .19476
                /gene="AT4g14460"
                /number=2
intron          19477. .19551
                /gene="AT4g14460"
                /number=2
exon            19552. .21256
                /gene="AT4g14460"
                /number=3
intron          21257. .21337
                /gene="AT4g14460"
                /number=3
exon            21338. .21472
                /gene="AT4g14460"
                /number=4
```

Query Match 19.7%; Score 191.6; DB 8; Length 197976;
Best Local Similarity 64.7%; Pred. No. 7.1e-43;
Matches 303; Conservative 0; Mismatches 159; Indels 6; Gaps 1;

```
Qy    161 cgtcgtccacgtggtcgtcctgcaggttccaaaaacaaacccaaacctccgacgattata 220
          ||| | ||| | ||| | ||  |||| ||||||||||||||||||| |||| |   || |
Db  24667 CGTAGACCAAGAGGTAGACCACCAGGATCCAAAAACAAACCCAAAGCTCCAATCTTTGTG 24726

Qy    221 actagagatagtcctaacgtccttagatcacacgttcttgaagtcacctccggttcggac 280
          || ||||| || || |||| ||  |    || || | ||  || ||  |||    |||
Db  24727 ACAAGAGACAGCCCCAACGCACTCCGTAGCCATGTCTTGGAGATCTCCGACGGCAGTGAC 24786

Qy    281 atatccgaggcagtctccacctacgccactcgtcgcggctgcggcgtttgcattataagc 340
            |  ||||  || || |   || | | |  ||  |     ||||||||||| || | |||
Db  24787 GTCGCCGACACAATCGCTCACTTCTCAAGACGCAGGCAACGCGGCGTTTGCGTTCTCAGC 24846

Qy    341 ggcacgggtgcggtcactaacgtcacgatacggcaacctgcggctccggctggtggaggt 400
          || || ||  | ||| |||||||||  | || ||| | || || ||       ||||||
Db  24847 GGGACAGGCTCAGTCGCTAACGTCACCCTCCGCCAAGCCGCCGCACC------AGGAGGT 24900

Qy    401 gtgattaccctgcatggtcggtttgacattttgtctttgaccggtactgcgcttccaccg 460
          ||| |  | || || ||  ||||||| || || ||||| |||||| ||   || ||
Db  24901 GTGGTCTCTCTCCAAGGCAGGTTTGAAATCTTATCTTTAACCGGTGCTTTCCTCCCTGGA 24960

Qy    461 cctgcaccaccgggagcaggaggtttgacggtgtatctagccggaggtcaaggacaagtt 520
          ||| | ||||| ||  ||   ||||| ||||| ||  |||||||  |  || || || ||
Db  24961 CCTTCCCCACCCGGGTCAACCGGTTTAACGGTTTACTTAGCCGGGGTCCAGGGTCAGGTC 25020

Qy    521 gtaggagggaatgtggctggttcgttaattgcttcgggaccggtagtgttgatggctgct 580
          || ||||| |  || |  ||  |  |  | ||    ||  ||||  || |||| ||||||
Db  25021 GTTGGAGGTAGCGTTGTAGGCCCACTCTTAGCCATAGGGTCGGTCATGGTGATTGCTGCT 25080

Qy    581 tcttttgcaaacgcagtttatgataggttaccgattgaagaggaagaa 628
          ||||  | |||||   |||||| || || || || ||||| || |||
Db  25081 ACTTTCTCTAACGCTACTTATGAGAGATTGCCCATGGAAGAAGAGGAA 25128
```

```
RESULT  11
ATFCA1
LOCUS       ATFCA1                206606 bp    DNA     linear   PLN 29-JUN-1999
```

```
DEFINITION  Arabidopsis thaliana DNA chromosome 4, ESSA I FCA contig fragment
            No. 1.
ACCESSION   Z97336
VERSION     Z97336.1  GI:2244788
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 206606)
  AUTHORS   Bevan,M., Stiekema,W., Murphy,G., Wambutt,R., Pohl,T., Terryn,N.,
            Kreis,M., Kavanagh,T., Entian,K.D., Rieger,M., James,R.,
            Puigdomenech,P., Hatzopoulos,P., Obermaier,B., Duesterhoft,A.,
            Jones,J., Palme,K., Ansorge,W., Delseny,M., Bancroft,I.,
            Mewes,H.W., Schueller,C. and Chalwatzis,N.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 206606)
  AUTHORS   EU Arabidopsis sequencing,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUN-1999) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            schuelle@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
COMMENT     Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
            and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
            this fragment has an overlap with ATFCA0 at the 5' end and an
            overlap with ATFCA2 at the 3' end.
FEATURES             Location/Qualifiers
     source          1. .206606
                     /organism="Arabidopsis thaliana"
                     /variety="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="4"
     source          1. .26121
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /clone="BAC TAMU8H22"
     gene            complement(join(685. .744,829. .903,1053. .1232,1804. .1806))
                     /gene="dl3200c"
     gene            685. .1806
                     /gene="dl3200c"
     exon            complement(685. .744)
                     /gene="dl3200c"
                     /number=1
     CDS             complement(join(685. .744,829. .903,1053. .1232,1804. .1806))
                     /gene="dl3200c"
                     /note="strong similarity to ribosomal protein L41, Candida
                     maltosa"
                     /codon_start=1
                     /product="ribosomal protein"
                     /protein_id="CAB10211.1"
                     /db_xref="GI:2244789"
                     /db_xref="SWISS-PROT:O23290"
                     /translation="MVNIPKTKNTYCKNKECKKHTLHKVTQYKKGKDSLAAQGKRRYD
                     RKQSGYGGQTKPVFHKKAKTTKKIVLRLQCQSCKHFSQRPIKRCKHFEIGGDKKGKGT
                     SLF"
     intron          complement(745. .828)
                     /gene="dl3200c"
                     /number=1
     exon            complement(829. .903)
                     /gene="dl3200c"
                     /number=2
     intron          complement(904. .1052)
                     /gene="dl3200c"
                     /number=2
     exon            complement(1053. .1232)
                     /gene="dl3200c"
                     /number=3
     intron          complement(1233. .1803)
                     /gene="dl3200c"
                     /number=3
     exon            complement(1804. .1806)
                     /gene="dl3200c"
                     /number=4
     exon            2229. .2837
                     /gene="dl3205w"
                     /number=1
     gene            2229. .7384
                     /gene="dl3205w"
     CDS             join(2229. .2837,2914. .3069,3161. .3268,3358. .3468,
                     3569. .4957,5000. .5212,6340. .6428,6469. .6560,6706. .6749,
                     7316. .7384)
                     /gene="dl3205w"
                     /note="similarity to kinesin osm-3, Caenorhabditis
                     elegans
                     contains EST gb:496806"
                     /codon_start=1
                     /product="kinesin like protein"
                     /protein_id="CAB10212.1"
                     /db_xref="GI:2244790"
                     /db_xref="SPTREMBL:O23291"
                     /translation="MAPTPSSSRSNQTQYTLIRTPQTKQRLNFHSKTPNPDGSKDPSP
                     PEHPVEVIGRIRDYPDRKEKSPSILQVNTDNQTVRVRADVGYRDFTLDGVSFSEQEGL
                     EEFYKKFIEERIKGVKVGNKCTIMMYGPTGAGKSHTMFGCGKEPGIVYRSLRDILGDS
                     DQDGVTFVQVTVLEVYNEEIYDLLSTNSSNNLGIGWPKGASTKVRLEVMGKKAKNASF
                     ISGTEAGKISKEIVKVEKRRIVKSTLCNERSSRSHCIIILDVPTVGGRLMLVDMAGSE
                     NIDQAGQTGFEAKMQTAKINQGNIALKRVVESIANGDSHVPFRDSKLTMLLQDSFEDD
                     KSKILMILCASPDPKEMHKTLCTLEYGAKAKCIVRGSHTPNKDKYGGDESASAVILGS
                     RIAAMDEFIIKLQSEKKQKEKERNEAQKQLKKKEEEVAALRSLLTQREACATNEEEIK
                     EKVNERTQLLKSELDKKLEECRRMAEEFVEMERRRMEERIVQQQEELEMMRRRLEEIE
                     VEFRRSNGGSVDETSGFAKRLRSLYSDDDPGMVKSMDLDMGDPEPVKQVWGAVSHQSS
                     NTISSNFTNLLQPKPSENMLTQMYPDRVCLSTVFEEEEVEEEEEKVIVEDKSICLITT
                     PMPSLNSEGLGKENCFNGADDKESASSRRLRIQNIFTLCGNQRELSQHSGQEEDQANI
                     ASPDKKDNQFFSITNKAEALAVEEAKENNISVDQRENGQLDIYVKWETAADNPRKLIT
                     TLRVTKDATLADLRKLIEIYLGSDNQAFTFLKLGVINLNQQAQKAFHFYLFEPCGAQV
                     AKEKESTVQATSLPLCNGHAYLATLRPGKSSQHKSLQPASPLPLNPIENMMEVTPISK
                     VTPNHQILLKTETERLGEADYINSLSLSHFLILFPRGESHQHRMLSFHSHQTSPSLSS
                     FPLLSRADADEPGLVLDITPLFEVVLEGPTLVLELAVVNDRHIAG"
     intron          2838. .2913
                     /gene="dl3205w"
                     /number=1
     exon            2914. .3069
                     /gene="dl3205w"
                     /number=2
     intron          3070. .3160
                     /gene="dl3205w"
                     /number=2
     exon            3161. .3268
                     /gene="dl3205w"
                     /number=3
     intron          3269. .3357
                     /gene="dl3205w"
                     /number=3
     exon            3358. .3468
                     /gene="dl3205w"
                     /number=4
     intron          3469. .3568
                     /gene="dl3205w"
                     /number=4
     exon            3569. .4957
                     /gene="dl3205w"
                     /number=5
     intron          4958. .4999
                     /gene="dl3205w"
                     /number=5
     exon            5000. .5212
                     /gene="dl3205w"
                     /number=6
     intron          5213. .6339
                     /gene="dl3205w"
                     /number=6
     exon            6340. .6428
```

```
                     /gene="dl3205w"
                     /number=7
     intron          6429. .6468
                     /gene="dl3205w"
                     /number=7
     exon            6469. .6560
                     /gene="dl3205w"
                     /number=8
     gene            complement(join(6533. .6772,6888. .7024,7289. .7352,
                     7451. .7524,7658. .7784,7870. .7954,8046. .8109,8216. .8277,
                     8489. .8559,8688. .8783,8860. .9008,9099. .9168,9435. .9475,
                     9576. .9669))
                     /gene="dl3210c"
     gene            6533. .9669
                     /gene="dl3210c"
     CDS             complement(join(6533. .6772,6888. .7024,7289. .7352,
                     7451. .7524,7658. .7784,7870. .7954,8046. .8109,8216. .8277,
                     8489. .8559,8688. .8783,8860. .9008,9099. .9168,9435. .9475,
                     9576. .9669))
                     /gene="dl3210c"
                     /note="CKI1 mRNA for casein kinase I
                     contains EST gb:957978;1054381"
                     /codon_start=1
                     /product="casein kinase I"
                     /protein_id="CAB10213.1"
                     /db_xref="GI:2244791"
                     /db_xref="SPTREMBL:Q39050"
                     /translation="MDRNQKMDHVIGGKFKLGRKLGSGSFGELYLGINIQTGEEVAVK
                     LEPVKTRHPQLQYESKIYMFLQGGTGVPHLKWFGVEGEYSCMVIDLLGPSLEDLFNYC
                     KRIFSLKSVLMLADQLICRVEYMHSRGFLHRDIKPDNFLMGLGRRANQVYIIDYGLAK
                     KYKDLQTQKHIPYRENKNLTGTARYASVNTHLGIEQSRRDDLESLGYVLMYFLRGSLP
                     WQGLKAGTKKQKYDKISEKKMLTSVETLCKSYPSEFTSYFHYCRSLRFEDKPDYSYLR
                     RLFRDLFIREGYQLDYVFDWTISKYPQIGSSSRPRPTPRPALDPPGPPAERAEKPTVG
                     QDLRGRFTGAIEAFTRRNVSSQGALGDRSRHRSSDDIPSSAKEVHESRNGSTSKRGVI
                     SSTRPGSSAEPSENHSSRLFSSGSRHATTQRVPQSYESAAAARPGHEDAIRNFELLTI
                     GSGKKRK"
     exon            complement(6533. .6772)
                     /gene="dl3210c"
                     /number=1
     intron          6561. .6705
                     /gene="dl3210c"
                     /number=8
     exon            6706. .6749
                     /gene="dl3210c"
                     /number=9
     intron          6750. .7315
                     /gene="dl3210c"
                     /number=9
     intron          complement(6773. .6887)
                     /gene="dl3210c"
                     /number=1
     exon            complement(6888. .7024)
                     /gene="dl3210c"

  Query Match             19.7%;  Score 191.6;  DB 8;  Length 206606;
  Best Local Similarity   64.7%;  Pred. No. 7.1e-43;
  Matches  303;  Conservative    0;  Mismatches  159;  Indels    6;  Gaps    1;

Qy     161 cgtcgtccacgtggtcgtcctgcaggttccaaaaacaaacccaaacctccgacgattata 220
           ||| | ||| | ||| | ||  |||| |||||||||||||||||| |||| |   || |
Db   79171 CGTAGACCAAGAGGTAGACCACCAGGATCCAAAAACAAACCCAAAGCTCCAATCTTTGTG 79230

Qy     221 actagagatagtcctaacgtccttagatcacacgttcttgaagtcacctccggttcggac 280
           || ||||| || || ||||  ||  |    || ||  | ||  || ||  |||    |||
Db   79231 ACAAGAGACAGCCCCAACGCACTCCGTAGCCATGTCTTGGAGATCTCCGACGGCAGTGAC 79290

Qy     281 atatccgaggcagtctccacctacgccactcgtcgcggctgcggcgtttgcattataagc 340
           |  ||||  || || |   || | | |  ||  |     |||||||||| || | |||
Db   79291 GTCGCCGACACAATCGCTCACTTCTCAAGACGCAGGCAACGCGGCGTTTGCGTTCTCAGC 79350

Qy     341 ggcacgggtgcggtcactaacgtcacgatacggcaacctgcggctccggctggtggaggt 400
           || || ||  | ||| |||||||||  | || ||| | || || ||       ||||||
Db   79351 GGGACAGGCTCAGTCGCTAACGTCACCCTCCGCCAAGCCGCCGCACC------AGGAGGT 79404

Qy     401 gtgattaccctgcatggtcggtttgacattttgtctttgaccggtactgcgcttccaccg 460
           ||| |  | || || ||  ||||||| || || ||||| |||||| ||   || ||
Db   79405 GTGGTCTCTCTCCAAGGCAGGTTTGAAATCTTATCTTTAACCGGTGCTTTCCTCCCTGGA 79464

Qy     461 cctgcaccaccgggagcaggaggtttgacggtgtatctagccggaggtcaaggacaagtt 520
           ||| | ||||| ||  ||   ||||| ||||| ||  |||||||  |  || || || ||
Db   79465 CCTTCCCCACCCGGGTCAACCGGTTTAACGGTTTACTTAGCCGGGGTCCAGGGTCAGGTC 79524

Qy     521 gtaggagggaatgtggctggttcgttaattgcttcgggaccggtagtgttgatggctgct 580
           || ||||| |  || |  ||  |  |  | ||    ||  ||||  || |||| ||||||
Db   79525 GTTGGAGGTAGCGTTGTAGGCCCACTCTTAGCCATAGGGTCGGTCATGGTGATTGCTGCT 79584

Qy     581 tcttttgcaaacgcagtttatgataggttaccgattgaagaggaagaa 628
           ||||  | |||||  |||||| || || || || ||||| || |||
Db   79585 ACTTTCTCTAACGCTACTTATGAGAGATTGCCCATGGAAGAAGAGGAA 79632


RESULT 12
AF446359
LOCUS       AF446359               1161 bp    mRNA    linear   PLN 19-DEC-2001
DEFINITION  Arabidopsis thaliana AT4g17800/dl4935c mRNA, complete cds.
ACCESSION   AF446359
VERSION     AF446359.1  GI:17933298
KEYWORDS    FLI_CDNA.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1161)
  AUTHORS   Shinn,P., Chen,H., Cheuk,R., Kim,C.J., Koesema,E., Meyers,M.C.,
            Banh,J., Bowser,L., Carninci,P., Chang,E., Dale,J.M.,
            Goldsmith,A.D., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A.,
            Karlin-Neumann,G., Kawai,J., Lam,B., Lee,J.M., Lin,J., Miranda,M.,
            Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L.,
            Sakurai,T., Satou,M., Seki,M., Southwick,A., Tang,C.C., Toriumi,M.,
            Wu,H.C., Yamada,K., Yamamura,Y., Yu,G., Yu,S., Shinozaki,K.,
            Davis,R.W., Theologis,A. and Ecker,J.R.
  TITLE     Arabidopsis cDNA clones
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1161)
  AUTHORS   Shinn,P., Chen,H., Cheuk,R., Kim,C.J., Koesema,E., Meyers,M.C.,
            Banh,J., Bowser,L., Carninci,P., Chang,E., Dale,J.M.,
            Goldsmith,A.D., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A.,
            Karlin-Neumann,G., Kawai,J., Lam,B., Lee,J.M., Lin,J., Miranda,M.,
            Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L.,
            Sakurai,T., Satou,M., Seki,M., Southwick,A., Tang,C.C., Toriumi,M.,
            Wu,H.C., Yamada,K., Yamamura,Y., Yu,G., Yu,S., Shinozaki,K.,
            Davis,R.W., Theologis,A. and Ecker,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-NOV-2001) Salk Institute Genomic Analysis Laboratory
            (SIGnAL), Plant Biology Laboratory, The Salk Institute for
            Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
            USA
COMMENT     RIKEN Genomic Sciences Center (GSC) members carried out the
            collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
            Arabidopsis Full-Length cDNA') :  Seki,M., Narusaka,M., Ishida,J.,
            Satou,M., Kamiya,A., Sakurai,T., Carninci,P.,  Kawai,J.,
            Hayashizaki,Y. and Shinozaki,K.

            The Salk, Stanford, PGEC (SSP) Consortium members carried out the
            sequencing and annotation of the RAFL cDNAs: Shinn,P., Chen,H.,
            Cheuk,R., Kim,C.J., Koesema,E., Meyers,M.C., Banh,J. Bowser,L.,
            Chang,E., Dale,J.M., Goldsmith,A.D., Jones,T., Karlin-Neumann,G.,
            Lam,B., Lee,J.M., Lin,J., Miranda,M., Nguyen,M., Onodera,C.S.,
            Palm,C.J., Quach,H.L., Southwick,A., Tang,C.C., Toriumi,M.,
            Wu,H.C., Yamada,K., Yamamura, Y., Yu,G., Yu,S., Davis,R.W.,
            Theologis,A., and Ecker,J.R.

            Shinn,P. (SSP/Salk) and Seki,M. (RIKEN GSC) contributed equally to
            this work.  Shinozaki,K. (RIKEN GSC) and Ecker,J.R. (SSP/Salk)
            contributed equally to this work as PIs.
```

```
FEATURES             Location/Qualifiers
     source          1. .1161
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /chromosome="4"
                     /clone="RAFL09-21-N24(R21145)"
                     /note="ecotype: Columbia"
     CDS             25. .903
                     /note="probable AT-hook DNA-binding protein"
                     /codon_start=1
                     /product="AT4g17800/dl4935c"
                     /protein_id="AAL48232.1"
                     /db_xref="GI:17933299"
                     /translation="MAGLDLGTAFRYVNHQLHRPDLHLHHNSSSDDVTPGAGMGHFTV
                     DDEDNNNNHQGLDLASGGGSGSSGGGGGHGGGGDVVGRRPRGRPPGSKNKPKPPVIIT
                     RESANTLRAHILEVTNGCDVFDCVATYARRRQRGICVLSGSGTVTNVSIRQPSAAGAV
                     VTLQGTFEILSLSGSFLPPPAPPGATSLTIFLAGGQGQVVGGSVVGELTAAGPVIVIA
                     ASFTNVAYERLPLEEDEQQQQLGGGSNGGGNLFPEVAAGGGGGLPFFNLPMNMQPNVQ
                     LPVEGWPGNSGGRGPF"
BASE COUNT       276 a    249 c    316 g    320 t
ORIGIN
```

```
  Query Match             19.5%;  Score 190.4;  DB 8;  Length 1161;
  Best Local Similarity   64.5%;  Pred. No. 9.3e-43;
  Matches  303;  Conservative    0;  Mismatches  161;  Indels    6;  Gaps    1;

Qy    158 ggacgtcgtccacgtggtcgtcctgcaggttccaaaaacaaacccaaacctccgacgatt 217
          || ||||||||||||||  | ||  | || ||||| |||||||| |||||||||   |||
Db    265 GGTCGTCGTCCACGTGGCAGACCACCGGGATCCAAGAACAAACCGAAACCTCCGGTAATT 324

Qy    218 ataactagagatagtcctaacgtccttagatcacacgttcttgaagtcacctccggttcg 277
          || ||  | || ||  | |||   || ||| | ||| |||||||||| ||   ||| |
Db    325 ATCACGCGCGAGAGCGCAAACACTCTAAGAGCTCACATTCTTGAAGTAACAAACGGCTGC 384

Qy    278 gacatatccgaggcagtctccacctacgccactcgtcgcggctgcggcgtttgcattata 337
          ||  | | |||    ||  | || || ||   |||  |     | ||  | ||| || |
Db    385 GATGTTTTCGACTGCGTTGCGACTTATGCTCGTCGGAGACAGCGAGGGATCTGCGTTCTG 444

Qy    338 agcggcacgggtgcggtcactaacgtcacgatacggcaacctgcggctccggctggtgga 397
          ||||| |  ||  |||||||| |||||||  ||||| ||      | |  | || | ||||
Db    445 AGCGGTAGCGGAACGGTCACGAACGTCAGCATACGTCA------GCCATCTGCGGCTGGA 498

Qy    398 ggtgtgattaccctgcatggtcggtttgacattttgtctttgaccggtactgcgcttcca 457
          |  ||  | || || || ||   ||| || ||| | ||| |  ||||  |    |||||
Db    499 GCGGTTGTGACGCTACAAGGAACGTTCGAGATTCTTTCTCTCTCCGGATCGTTTCTTCCT 558

Qy    458 ccgcctgcaccaccgggagcaggaggtttgacggtgtatctagccggaggtcaaggacaa 517
          || || ||||| || ||||||    |||||||  | |   ||||||||| ||||| ||
Db    559 CCTCCGGCACCTCCCGGAGCAACGAGTTTGACAATTTTCTTAGCCGGAGGACAAGGTCAG 618

Qy    518 gttgtaggagggaatgtggctggttcgttaattgcttcgggaccggtagtgttgatggct 577
          || || ||||| |  || |  |||  | | |  ||  | ||||||||  |  |||| ||
Db    619 GTGGTTGGAGGAAGCGTTGTGGGTGAGCTTACGGCGGCTGGACCGGTGATTGTGATTGCA 678

Qy    578 gcttctttttgcaaacgcagtttatgataggttaccgattgaagaggaaga 627
          ||||| ||| | || |  | |||||| ||  | ||  | ||||| || ||
Db    679 GCTTCGTTTACTAATGTTGCTTATGAGAGACTTCCTTTAGAAGAAGATGA 728
```

```
RESULT  13
ATFCA9/c
LOCUS       ATFCA9                  81835 bp    DNA     linear   PLN 28-JUN-1999
DEFINITION  Arabidopsis thaliana DNA chromosome 4, ESSA I FCA contig fragment
            No. 9.
ACCESSION   Z97344
VERSION     Z97344.2  GI:5281025
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 81835)
  AUTHORS   Bevan,M., Stiekema,W., Murphy,G., Wambutt,R., Pohl,T., Terryn,N.,
            Kreis,M., Kavanagh,T., Entian,K.D., Rieger,M., James,R.,
            Puigdomenech,P., Hatzopoulos,P., Obermaier,B., Duesterhoft,A.,
            Jones,J., Palme,K., Ansorge,W., Delseny,M., Bancroft,I.,
            Mewes,H.W., Schueller,C. and Chalwatzis,N.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 81835)
  AUTHORS   EU Arabidopsis sequencing,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUN-1999) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            schuelle@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
COMMENT     On Jun 30, 1999 this sequence version replaced gi:2245126.
            Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
            and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
            this fragment has an overlap with ATFCA8 at the 5' end.
FEATURES             Location/Qualifiers
     source          1. .81835
                     /organism="Arabidopsis thaliana"
                     /variety="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="4"
     source          1. .73596
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /clone="BAC TAMU3P6"
     gene            complement(join(8240. .8764,8837. .9087,9167. .9237,
                     9335. .9381))
                     /gene="dl4875c"
     gene            8240. .9381
                     /gene="dl4875c"
     CDS             complement(join(8240. .8764,8837. .9087,9167. .9237,
                     9335. .9381))
                     /gene="dl4875c"
                     /note="weak similarity to Mus musculus IAP(inhibit
                     apoptosis) homolog A (MIHA) mRNA"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAB10548.1"
                     /db_xref="GI:2245127"
                     /db_xref="SPTREMBL:O23608"
                     /translation="MAIQAQHPSRFFFNNSNGQEASDCSLQPQDTPFTNFTKAGVDSR
                     KRSREVYSSALMNPPPPKPSQVIDITELLQKTPNVVSTGLRLFHDQSQNQQQFFSSLP
                     GDVTGKIKRQRDELDRFIQTQGEELRRTLADNRERRYVELLCAAEEIVGRKLRKKEAE
                     LEKATRRHAELEARVAHIVEEARNWQLRAATREAEVSSLHAHLQQAIANRLDTAAKQS
                     TFGEDGGDAEEAEDAESVYVDPERIELIGPSCRICRRKSATVMALPCQHLILCNGCDV
                     GAVRVCPICLAVKTSGVEVLF"
     exon            complement(8240. .8764)
                     /gene="dl4875c"
                     /number=1
     intron          complement(8765. .8836)
                     /gene="dl4875c"
                     /number=1
     exon            complement(8837. .9087)
                     /gene="dl4875c"
                     /number=2
     intron          complement(9088. .9166)
                     /gene="dl4875c"
                     /number=2
     exon            complement(9167. .9237)
                     /gene="dl4875c"
                     /number=3
     intron          complement(9238. .9334)
                     /gene="dl4875c"
                     /number=3
     exon            complement(9335. .9381)
                     /gene="dl4875c"
                     /number=4
```

```
repeat_region   9966. .9988
                /note="23bp T tandem repeat"
gene            11464. .12444
                /gene="dl4880w"
CDS             11464. .12444
                /gene="dl4880w"
                /note="strong similarity to peroxidase (EC 1.11.1.7),
                Medicago sativa"
                /codon_start=1
                /product="peroxidase like protein"
                /protein_id="CAB10549.1"
                /db_xref="GI:2245128"
                /db_xref="SPTREMBL:O23609"
                /translation="MSSVINVLFVVLVFVPSIYSAPPPNLTKDYYQKTCPDFNKIVRE
                TVTPKQGQQPTTAAGTLRLFFHDCFMEGCDASVLIATNSFNKAERDDDLNESLPGDAF
                DIVTRIKTALELSCPGVVSCADILAQATRDLVTMVGGPFYEVKLGRKDGFESKAHKVK
                GNLPLANQSVPDMLSIFKKNGFTLKELVALSGGHTIGFSHCKEFSNRIFPKVDPELNA
                KFAGVLKDLCKNFETNKTMAAFLDPVTPGKFDNMYFKNLKRGLGLLASDHILFKDPST
                RPFVELYANNQTAFFEDFARAMEKLGRVGVKGEKDGEVRRRCDHFNKLNV"
exon            11464. .12444
                /gene="dl4880w"
                /number=1
repeat_region   15042. .15090
                /note="49bp TA tandem repeat"
repeat_region   17958. .17983
                /note="26bp AT tandem repeat"
gene            19659. .22020
                /gene="dl4885w"
CDS             join(19659. .19700,19880. .20074,20118. .20348,20653. .20742,
                20834. .20940,21627. .22020)
                /gene="dl4885w"
                /note="weak similarity to T08A11.2, Caenorhabditis
                elegans"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="CAB10550.1"
                /db_xref="GI:2245129"
                /db_xref="SPTREMBL:O23610"
                /translation="MEEYEFLVRVREAGGFDIEHLMDTKPSSCVIGCREFDNNDDYNA
                HIVLFARMGIHKYNMIQVKNLSPLVYICSSSSCNRGTNLQLSCVEKYNYRSPRVYSAF
                YITLIAKDPDACNSLVTFQTRVVEEGFNMTKLSCNIARPKPGPQGKGLGTLYQRVKKS
                ELLTHDWIRLYMELAFLKAHLRSPKNPNLSKLVIVMVAIETKDNMEPPSERLKARNAF
                FYINCDYLPKYPPNLLLITDVVDERGSDTVDSGANTVIDDGDGGKIESFGGDVAPQDL
                YCNSEQARRRTDGTALDTIGWIHWNNRESNGQKWYFRRIFAVHGELNSDGVDGRVHQE
                TPGVIGACFLKHQLYIMR"
exon            19659. .19700
                /gene="dl4885w"
                /number=1
intron          19701. .19879
                /gene="dl4885w"
                /number=1
exon            19880. .20074
                /gene="dl4885w"
                /number=2
intron          20075. .20117
                /gene="dl4885w"
                /number=2
exon            20118. .20348
                /gene="dl4885w"
                /number=3
intron          20349. .20652
                /gene="dl4885w"
                /number=3
exon            20653. .20742
                /gene="dl4885w"
                /number=4
intron          20743. .20833
                /gene="dl4885w"
                /number=4
exon            20834. .20940
                /gene="dl4885w"
                /number=5
intron          20941. .21626
                /gene="dl4885w"
                /number=5
exon            21627. .22020
                /gene="dl4885w"
                /number=6
exon            complement(21664. .22032)
                /gene="dl4890c"
                /number=1
gene            complement(join(21664. .22032,22287. .22522,22587. .22923,
                23083. .23181,23268. .23627,23704. .23856,24004. .24121,
                24229. .24392,24476. .24625))
                /gene="dl4890c"
gene            21664. .24625
                /gene="dl4890c"
CDS             complement(join(21664. .22032,22287. .22522,22587. .22923,
                23083. .23181,23268. .23627,23704. .23856,24004. .24121,
                24229. .24392,24476. .24625))
                /gene="dl4890c"
                /note="similarity to homeotic protein GL2, Arabidopsis
                thaliana"
                /codon_start=1
                /product="GLABRA2 like protein"
                /protein_id="CAB10551.1"
                /db_xref="GI:2245130"
                /db_xref="SPTREMBL:O23611"
                /translation="METKDKKEKGHMVLNSDNVFGSVSSSPTTTIQNPNYFTSFENPN
                FPYIFPKEEYEVMSKIESGSGKSTGSGHDPVENTAIEQEPPAAKKKRYHRHTASQIQQ
                MEALFKENAHPDTKTRLRLSKKLGLSPIQVKFWFQNKRTQIKAQQSRSDNAKLKAENE
                TLKTESQNIQSNFQCLFCSTCGHNLRLENARLRQELDRLRSIVSMRNPSPSQEITPET
                NKNNNDNMLIAEEEKAIDMELAVSCARELAKMCDINEPLWNKKRLDNESVCLNEEEYK
                KMFLWPLMNDDDRFRREASRANAVIMLNCITLVKAFLDADKWSEMFFPIVSSAKTAQI
                ISSGASGPSGTLLLQNAEEGKWMVVDFPIDRIKPASATTTDQYRRKPSGCIIQAMRNG
                YSQVTWVEHVEVEEEKHVQDEVVREFVESGVAFGAERWLSVLKRQCERMASLMATNITD
                LGGIYITCFTMIPSVEARKNLMKLSQRMVKTFCLNIINSHGQAPTKDTVKIVSRKVCG
                GLVPCAVSVTLLPYSHQQVFDLLRDNQRLSQVESNSSHNVELMLQETCTDNSGSLLVY
                STVDPVAVQLAMNGEDPSEIPLLPVGFSVVPVNPSDGVEGSSVSSPSCLLTVAIQVLG
                SNVTTERLDLSTVSVINHRICATVNRITSALVNDVGN"
intron          complement(22033. .22286)
                /gene="dl4890c"
                /number=1
exon            complement(22287. .22522)
                /gene="dl4890c"
                /number=2
```

Query Match 19.5%; Score 190.4; DB 8; Length 81835;
Best Local Similarity 64.5%; Pred. No. 1.4e-42;
Matches 303; Conservative 0; Mismatches 161; Indels 6; Gaps 1;

```
Qy    158 ggacgtcgtccacgtggtcgtcctgcaggttccaaaaacaaacccaaacctccgacgatt 217
          || ||||||||||||||  | ||  | || ||||| |||||||| |||||||||   |||
Db  61524 GGTCGTCGTCCACGTGGCAGACCACCGGGATCCAAGAACAAACCGAAACCTCCGGTAATT 61465

Qy    218 ataactagagatagtcctaacgtccttagatcacacgttcttgaagtcacctccggttcg 277
          || ||  | || ||  | |||   || ||| | ||| |||||||||| ||   ||| |
Db  61464 ATCACGCGCGAGAGCGCAAACACTCTAAGAGCTCACATTCTTGAAGTAACAAACGGCTGC 61405

Qy    278 gacatatccgaggcagtctccacctacgccactcgtcgcggctgcggcgtttgcattata 337
          ||  | | |||    ||  | || || ||   ||| |     | ||  | ||| || |
Db  61404 GATGTTTTCGACTGCGTTGCGACTTATGCTCGTCGGAGACAGCGAGGGATCTGCGTTCTG 61345

Qy    338 agcggcacgggtgcggtcactaacgtcacgatacggcaacctgcggctccggctggtgga 397
          ||||| |  ||  |||||||| |||||||  ||||| ||      | |  | || | ||||
Db  61344 AGCGGTAGCGGAACGGTCACGAACGTCAGCATACGTCA------GCCATCTGCGGCTGGA 61291

Qy    398 ggtgtgattaccctgcatggtcggtttgacattttgtctttgaccggtactgcgcttcca 457
          |  ||  | || || || ||   ||| || ||| | ||| |  ||||  |    |||||
Db  61290 GCGGTTGTGACGCTACAAGGAACGTTCGAGATTCTTTCTCTCTCCGGATCGTTTCTTCCT 61231

Qy    458 ccgcctgcaccaccgggagcaggaggtttgacggtgtatctagccggaggtcaaggacaa 517
          || || ||||| || ||||||    |||||||  | |   ||||||||| ||||| ||
Db  61230 CCTCCGGCACCTCCCGGAGCAACGAGTTTGACAATTTTCTTAGCCGGAGGACAAGGTCAG 61171

Qy    518 gttgtaggagggaatgtggctggttcgttaattgcttcgggaccggtagtgttgatggct 577
```

```
          || || ||||| |  || |  |||  | | |  ||  | ||||||||  |  |||| ||
Db  61170 GTGGTTGGAGGAAGCGTTGTGGGTGAGCTTACGGCGGCTGGACCGGTGATTGTGATTGCA 61111

Qy    578 gcttcttttgcaaacgcagtttatgataggttaccgattgaagaggaaga 627
          ||||| ||| | || |  | ||||| ||  | ||  | ||||| || ||
Db  61110 GCTTCGTTTACTAATGTTGCTTATGAGAGACTTCCTTTAGAAGAAGATGA 61061


RESULT  14
ATCHRIV47/c
LOCUS       ATCHRIV47              198067 bp    DNA     linear   PLN 16-MAR-2000
DEFINITION  Arabidopsis thaliana DNA chromosome 4, contig fragment No. 47.
ACCESSION   AL161547
VERSION     AL161547.2  GI:7268521
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 77203 to 176024)
  AUTHORS   Rieger,M., Mueller-Auer,S., Zipp,M., Schaefer,M., Mewes,H.W.,
            Lemcke,K. and Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   2  (bases 173478 to 198067)
  AUTHORS   Murphy,G., Ridley,P., Hudson,S., Mewes,H.W., Lemcke,K. and
            Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   3  (bases 76382 to 77642)
  AUTHORS   Robben,J., Grymonprez,B., Volckaert,G., Mewes,H.W., Lemcke,K. and
            Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   4  (bases 1 to 198067)
  AUTHORS   EU Arabidopsis sequencing,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
COMMENT     Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
            and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
            this fragment has an overlap with ATCHRIV46 at the 5' end and an
            overlap with ATCHRIV48 at the 3' end.
FEATURES             Location/Qualifiers
     source          1. .198067
                     /organism="Arabidopsis thaliana"
                     /variety="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="4"
     gene            6476. .7456
                     /gene="AT4g17690"
     CDS             6476. .7456
                     /gene="AT4g17690"
                     /note="strong similarity to peroxidase (EC 1.11.1.7),
                     Medicago sativa
                     Contains Peroxidases signatures AA186-196;Peroxidases
                     signatures AA57-68
                     contains EST gb:AI999391.1, AI994341.1, AI994675.1,
                     AI995254.1, AA721954, R90036, Z35219, AA713329"
                     /codon_start=1
                     /product="peroxidase like protein"
                     /protein_id="CAB78772.1"
                     /db_xref="GI:7268522"
                     /translation="MSSVINVLFVVLVFVPSIYSAPPPNLTKDYYQKTCPDFNKIVRE
                     TVTPKQGQQPTTAAGTLRLFFHDCFMEGCDASVLIATNSFNKAERDDDLNESLPGDAF
                     DIVTRIKTALELSCPGVVSCADILAQATRDLVTMVGGPFYEVKLGRKDGFESKAHKVK
                     GNLPLANQSVPDMLSIFKKNGFTLKELVALSGGHTIGFSHCKEFSNRIFPKVDPELNA
                     KFAGVLKDLCKNFETNKTMAAFLDPVTPGKFDNMYFKNLKRGLGLLASDHILFKDPST
                     RPFVELYANNQTAFFEDFARAMEKLGRVGVKGEKDGEVRRRCDHFNKLNV"
     exon            6476. .7456
                     /gene="AT4g17690"
                     /number=1
repeat_region        10054. .10102
repeat_region        12970. .12995
exon                 14671. .14712
                     /gene="AT4g17700"
                     /number=1
gene                 14671. .17032
                     /gene="AT4g17700"
CDS                  join(14671. .14712,14892. .15086,15130. .15360,15665. .15754,
                     15846. .15952,16639. .17032)
                     /gene="AT4g17700"
                     /note="similarity to T08A11.2, Caenorhabditis elegans
                     Contains Prokaryotic membrane lipoprotein lipid attachment
                     site AA67-77"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAB78773.1"
                     /db_xref="GI:7268523"
                     /translation="MEEYEFLVRVREAGGFDIEHLMDTKPSSCVIGCREFDNNDDYNA
                     HIVLFARMGIHKYNMIQVKNLSPLVYICSSSSCNRGTNLQLSCVEKYNYRSPRVYSAF
                     YITLIAKDPDACNSLVTFQTRVVEEGFNMTKLSCNIARPKPGPQGKGLGTLYQRVKKS
                     ELLTHDWIRLYMELAFLKAHLRSPKNPNLSKLVIVMVAIETKDNMEPPSERLKARNAF
                     FYINCDYLPKYPPNLLLITDVVDERGSDTVDSGANTVIDDGDGGKIESFGGDVAPQDL
                     YCNSEQARRRTDGTALDTIGWIHWNNRESNGQKWYFRRIFAVHGELNSDGVDGRVHQE
                     TPGVIGACFLKHQLYIMR"
intron               14713. .14891
                     /gene="AT4g17700"
                     /number=1
exon                 14892. .15086
                     /gene="AT4g17700"
                     /number=2
intron               15087. .15129
                     /gene="AT4g17700"
                     /number=2
exon                 15130. .15360
                     /gene="AT4g17700"
                     /number=3
intron               15361. .15664
                     /gene="AT4g17700"
                     /number=3
exon                 15665. .15754
                     /gene="AT4g17700"
                     /number=4
intron               15755. .15845
                     /gene="AT4g17700"
                     /number=4
exon                 15846. .15952
                     /gene="AT4g17700"
                     /number=5
intron               15953. .16638
                     /gene="AT4g17700"
                     /number=5
exon                 16639. .17032
                     /gene="AT4g17700"
                     /number=6
gene                 complement(join(16676. .17044,17299. .17534,17599. .17935,
                     18095. .18193,18280. .18639,18716. .18868,19016. .19133,
                     19241. .19404,19488. .19637))
                     /gene="AT4g17710"
CDS                  complement(join(16676. .17044,17299. .17534,17599. .17935,
                     18095. .18193,18280. .18639,18716. .18868,19016. .19133,
                     19241. .19404,19488. .19637))
                     /gene="AT4g17710"
                     /note="similarity to homeotic protein GL2, Arabidopsis
                     thaliana
                     Contains 'Homeobox' domain signature and profile
                     AA121-144"
                     /codon_start=1
                     /product="GLABRA2 like protein"
                     /protein_id="CAB78774.1"
                     /db_xref="GI:7268524"
                     /translation="METKDKKEKGHMVLNSDNVFGSVSSSPTTTIQNPNYFTSFENPN
```

```
                     FPYIFPKEEYEVMSKIESGSGKSTGSGHDPVENTAIEQEPPAAKKKRYHRHTASQIQQ
                     MEALFKENAHPDTKTRLRLSKKLGLSPIQVKFWFQNKRTQIKAQQSRSDNAKLKAENE
                     TLKTESQNIQSNFQCLFCSTCGHNLRLENARLRQELDRLRSIVSMRNPSPSQEITPET
                     NKNNNDNMLIAEEEKAIDMELAVSCARELAKMCDINEPLWNKKRLDNESVCLNEEEYK
                     KMFLWPLMNDDDRFRREASRANAVIMLNCITLVKAFLDADKWSEMFFPIVSSAKTAQI
                     ISSGASGPSGTLLLQNAEEGKWMVVDFPIDRIKPASATTTDQYRRKPSGCIIQAMRNG
                     YSQVTWVEHVEVEEKHVQDEVVREFVESGVAFGAERWLSVLKRQCERMASLMATNITD
                     LGGIYITCFTMIPSVEARKNLMKLSQRMVKTFCLNIINSHGQAPTKDTVKIVSRKVCG
                     GLVPCAVSVTLLPYSHQQVFDLLRDNQRLSQVESNSSHNVELMLQETCTDNSGSLLVY
                     STVDPVAVQLAMNGEDPSEIPLLPVGFSVVPVNPSDGVEGSSVSSPSCLLTVAIQVLG
                     SNVTTERLDLSTVSVINHRICATVNRITSALVNDVGN"
     exon            complement(16676. .17044)
                     /gene="AT4g17710"
                     /number=1
     gene            16676. .19637
                     /gene="AT4g17710"
     intron          complement(17045. .17298)
                     /gene="AT4g17710"
                     /number=1
     exon            complement(17299. .17534)
                     /gene="AT4g17710"
                     /number=2
     intron          complement(17535. .17598)
                     /gene="AT4g17710"
                     /number=2
     exon            complement(17599. .17935)
                     /gene="AT4g17710"
                     /number=3
     intron          complement(17936. .18094)
                     /gene="AT4g17710"
                     /number=3
     exon            complement(18095. .18193)
                     /gene="AT4g17710"
                     /number=4
     intron          complement(18194. .18279)
                     /gene="AT4g17710"
                     /number=4
     exon            complement(18280. .18639)
                     /gene="AT4g17710"
                     /number=5
     intron          complement(18640. .18715)
                     /gene="AT4g17710"
                     /number=5
     exon            complement(18716. .18868)
                     /gene="AT4g17710"
                     /number=6
     intron          complement(18869. .19015)
                     /gene="AT4g17710"
                     /number=6
     exon            complement(19016. .19133)
                     /gene="AT4g17710"
                     /number=7
     intron          complement(19134. .19240)
                     /gene="AT4g17710"
                     /number=7
     exon            complement(19241. .19404)
                     /gene="AT4g17710"
                     /number=8
     intron          complement(19405. .19487)
                     /gene="AT4g17710"
                     /number=8
     exon            complement(19488. .19637)
                     /gene="AT4g17710"
                     /number=9

  Query Match             19.5%;  Score 190.4;  DB 8;  Length 198067;
  Best Local Similarity   64.5%;  Pred. No. 1.5e-42;
  Matches  303;  Conservative    0;  Mismatches  161;  Indels    6;  Gaps    1;

Qy    158 ggacgtcgtccacgtggtcgtcctgcaggttccaaaaacaaacccaaacctccgacgatt 217
          || ||||||||||||||  | ||  | || ||||| |||||||| |||||||||   |||
Db  56536 GGTCGTCGTCCACGTGGCAGACCACCGGGATCCAAGAACAAACCGAAACCTCCGGTAATT 56477

Qy    218 ataactagagatagtcctaacgtccttagatcacacgttcttgaagtcacctccggttcg 277
          || ||  | || ||  | |||   || ||| | ||| |||||||||| ||   ||| |
Db  56476 ATCACGCGCGAGAGCGCAAACACTCTAAGAGCTCACATTCTTGAAGTAACAAACGGCTGC 56417

Qy    278 gacatatccgaggcagtctccacctacgccactcgtcgcggctgcggcgtttgcattata 337
          ||  | | |||    ||  | || || ||   ||| |     | ||  | ||| || |
Db  56416 GATGTTTTCGACTGCGTTGCGACTTATGCTCGTCGGAGACAGCGAGGGATCTGCGTTCTG 56357

Qy    338 agcggcacgggtgcggtcactaacgtcacgatacggcaacctgcggctccggctggtgga 397
          ||||| |  ||  ||||||| |||||||  ||||| ||      | |  | || | ||||
Db  56356 AGCGGTAGCGGAACGGTCACGAACGTCAGCATACGTCA------GCCATCTGCGGCTGGA 56303

Qy    398 ggtgtgattaccctgcatggtcggtttgacattttgtctttgaccggtactgcgcttcca 457
          |  ||  | || || || ||   ||| || ||| | ||| |  |||  |   |||||
Db  56302 GCGGTTGTGACGCTACAAGGAACGTTCGAGATTCTTTCTCTCTCCGGATCGTTTCTTCCT 56243

Qy    458 ccgcctgcaccaccgggagcaggaggtttgacggtgtatctagccggaggtcaaggacaa 517
          || || ||||| || |||||||    |||||||  | |   |||||||||| ||||| ||
Db  56242 CCTCCGGCACCTCCCGGAGCAACGAGTTTGACAATTTTCTTAGCCGGAGGACAAGGTCAG 56183

Qy    518 gttgtaggagggaatgtggctggttcgttaattgcttcgggaccggtagtgttgatggct 577
          || || ||||| |  || |  |||  | | |  ||  | ||||||||  |  |||| ||
Db  56182 GTGGTTGGAGGAAGCGTTGTGGGTGAGCTTACGGCGGCTGGACCGGTGATTGTGATTGCA 56123

Qy    578 gcttctttttgcaaacgcagtttatgataggttaccgattgaagaggaaga 627
          ||||| ||| | || |  | |||||| ||  | ||  | ||||| || ||
Db  56122 GCTTCGTTTACTAATGTTGCTTATGAGAGACTTCCTTTAGAAGAAGATGA 56073


RESULT  15
ATT12H17
LOCUS       ATT12H17               89350 bp    DNA     linear   PLN 03-FEB-1998
DEFINITION  Arabidopsis thaliana DNA chromosome 4, BAC clone T12H17 (ESSAII
            project).
ACCESSION   AL021635
VERSION     AL021635.1  GI:2827538
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 89350)
  AUTHORS   Bevan,M., Hilbert,H., Braun,M., Holzer,E., Brandt,A.,
            Duesterhoeft,A., Bancroft,I., Mewes,H.W., Mayer,K. and Schueller,C.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 89350)
  AUTHORS   EU Arabidopsis sequencing,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-FEB-1998) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, Project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
FEATURES             Location/Qualifiers
     source          1. .89350
                     /organism="Arabidopsis thaliana"
                     /variety="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="4"
     gene            669. .1151
                     /gene="T12H17.10"
     exon            669. .1151
                     /gene="T12H17.10"
                     /number=1
     CDS             669. .1151
                     /gene="T12H17.10"
                     /note="similarity to auxin-induced protein 10A, Glycine
                     max., PIR2:JQ1099"
                     /codon_start=1
                     /product="predicted protein"
                     /protein_id="CAA16547.1"
                     /db_xref="GI:2827539"
```

```
                     /db_xref="SPTREMBL:O49643"
                     /translation="MKRLRGFKIGHRFVKIFKWIIRSRRNQTGKRQCLTGILNPVTKI
                     YSLARRCLRRGANRLCGGMKPGQTRLGNEPKTPTVPRGHLVVHVGESGEDTRRVVVPV
                     IYFNHPLFGELLEQAERVYGFEQPGRIMIPCRVSDFEKVQMRIAAWDHCRRKSTFKIL
                     "
     exon            2406. .2754
                     /gene="T12H17.20"
                     /number=1
     gene            2406. .2983
                     /gene="T12H17.20"
     CDS             join(2406. .2754,2844. .2983)
                     /gene="T12H17.20"
                     /note="similarity to predicted proteins, Arabidopsis
                     thaliana"
                     /codon_start=1
                     /product="predicted protein"
                     /protein_id="CAA16548.1"
                     /db_xref="GI:2827540"
                     /db_xref="SPTREMBL:O49644"
                     /translation="MAYTNKVAVAVGAAVVFLAVVMNPRWTEAQTYPKLDRLCVMMIP
                     DILEECFTHDRLKPTEDCCNDLKNATMTQVDCLCDNFLESLSFSDLSRTFSAGVLKKC
                     DVSHKYMCQAAKNRGEAKGGRNSTTTCDNSITNTSVGGKNKVATSMSAFGLVAILLFV
                     MF"
     intron          2755. .2843
                     /gene="T12H17.20"
                     /number=1
     exon            2844. .2983
                     /gene="T12H17.20"
                     /number=2
     tRNA            3327. .3400
                     /note="Asn-tRNA_(GTT)"
                     /product="tRNA-Asn"
     gene            4696. .5489
                     /gene="T12H17.30"
     CDS             join(4696. .5035,5407. .5489)
                     /gene="T12H17.30"
                     /note="contains EST gb:N38564"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAA16549.1"
                     /db_xref="GI:2827541"
                     /db_xref="SPTREMBL:O49645"
                     /translation="MAYTNKVTISAAVATMMLFLAVTIVDAQSMPPMPKFNPVCALAD
                     LPNIVQLCYFNLDLTPSEECCNDLKSSSTIQVNCLCDNFIAHPSNGNISQARYDLVNS
                     ACGVADKFACKGGDASGGSTNKIAASMVLLGLVASLFF"
     exon            4696. .5035
                     /gene="T12H17.30"
                     /number=1
     intron          5036. .5406
                     /gene="T12H17.30"
                     /number=1
     exon            5407. .5489
                     /gene="T12H17.30"
                     /number=2
     CDS             join(6595. .6940,7035. .7105)
                     /gene="T12H17.40"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAA16550.1"
                     /db_xref="GI:2827542"
                     /db_xref="SPTREMBL:O49646"
                     /translation="MAVAVTAVLFLAVVIAPQWTETKKPPRPSDTSDTSGTSGRDRRT
                     MCPLSIPGIVQNCYATLNAFPSKECCKDLKTASKREVTCLCNNVIAHPDPLYTNTNQV
                     NKACGVLDKYACDVAPTFTIVYTNYDMIIIAVIQDP"
     exon            6595. .6940
                     /gene="T12H17.40"
                     /number=1
     gene            6595. .7105
                     /gene="T12H17.40"
     intron          6941. .7034
                     /gene="T12H17.40"
                     /number=1
     exon            7035. .7105
                     /gene="T12H17.40"
                     /number=2
     exon            complement(8019. .9209)
                     /gene="T12H17.50"
                     /number=1
     gene            8019. .9209
                     /gene="T12H17.50"
     gene            complement(8019. .9209)
                     /gene="T12H17.50"
     CDS             complement(8019. .9209)
                     /gene="T12H17.50"
                     /note="similarity to predicted protein, Arabidopsis
                     thaliana"
                     /codon_start=1
                     /product="predicted protein"
                     /protein_id="CAA16551.1"
                     /db_xref="GI:2827543"
                     /db_xref="SPTREMBL:O49647"
                     /translation="MEKNQNPNTWSDLPLDLLNLVFKRLSFANFRQAKSVCSSWYSAS
                     KQSVPKNQIPWLMLFPKDKNNNKNSSCTIFFNPEDKDQLYQTQDLGVEFAKSVCLATY
                     GSWLLMQDSKYNLYILNPFTYEKIGLPAIESQQVGMVKVDQTIDDDFLTFDDHNHVKL
                     FKGNNTVRTPVFWIDEKTKEYIALWGLGYWCVVYAKNGDKLWNQIPEIILDSLDMVYK
                     DHKLYSFSYRNLFTILDFSGEIPRKAFQCFMHVYRSEWLSPRSRQLSNSWCVAETKLV
                     VTVTGDVLLVERMLRHWSRIQSFNVYKFYSSGTFLDKYELADSLGDEAMLLDLGITVL
                     ANEVEGLHRNTIYFSDSHDTTTKDLFLFNYETREMEPLHKFDFDCSLLELSARWFLPS
                     FSHT"
     gene            9990. .13709
                     /gene="T12H17.60"
     CDS             join(9990. .10023,10239. .10535,10564. .10664,10910. .11292,
                     11755. .11809,11898. .12053,12143. .12311,12391. .12455,
                     12534. .12674,12769. .12885,13024. .13096,13244. .13380,
                     13468. .13525,13612. .13709)
                     /gene="T12H17.60"
                     /note="similarity to p48, Homo sapiens, PATCHX:G904032
                     contains EST gb:Aa605571, H76919, H76638, R30271, Z25650"
                     /codon_start=1
                     /product="HSP associated protein like"
                     /protein_id="CAA16552.1"
                     /db_xref="GI:2827544"
                     /db_xref="SPTREMBL:O49648"
                     /translation="MRRTWVLAEESSVAIAPLLAEPQSTMFPEMTPECATVMPDLLEK
                     CFATGSVTPTEDCCTDLKSATSTQVTCLCDNYIANPAVSNITGPYSKAITTKCGVFDK
                     YSCDGTSKVISNKNGAFMVINGFWFCVVKEEKKRKEEAAVAMGVLTLLGLINGLIVAH
                     GKTRQQNKPVCKLILFQFQHFGFCSIRFSANKKARSRSKTGDAYPFRNEKVLEQKPYS
                     LPFVPQILEIETRLKTKKTRLLTLRPKPTLFLFLSLSRRSEADFSQKDGFNEAYLGAK
                     IPTGVHEEDKDTKPRSFVVEESDDDMDETEEVKPKVEEEEEEDEIVESDVELEGDTVE
                     PDNDPPQKMGDSSVEVTDENREAAQEAKGKAMEALSEGNFDEAIEHLTRAITLNPTSA
                     IMYGNRASVYIKLKKPNAAIRDANAALEINPDSAKGYKSRGMARAMLGEWAEAAKDLH
                     LASTIDYDEEISAVLKKVEPNAHKLEEHRRKYDRLRKEREDKKAERDRLRRRAEAQAA
                     YDKAKKEEQSSSSRPSGGGFPGGMGGGMPGMGGGMPAGMGGGGMPGAGGGMPGGGGMP
                     GGMDFSKILNDPELMTAFSDPEVMAALQDVMKNPANLAKHQANPKVAPVIAKMMGKFA
                     GPQ"
     exon            9990. .10023
                     /gene="T12H17.60"
                     /number=1
     intron          10024. .10238
                     /gene="T12H17.60"
                     /number=1
     exon            10239. .10535
                     /gene="T12H17.60"
                     /number=2
     intron          10536. .10563
                     /gene="T12H17.60"
                     /number=2
     exon            10564. .10664
                     /gene="T12H17.60"
                     /number=3
     intron          10665. .10909
                     /gene="T12H17.60"
                     /number=3
     exon            10910. .11292
                     /gene="T12H17.60"
                     /number=4
```

```
     intron          11293. .11754
                     /gene="T12H17.60"
                     /number=4
     exon            11755. .11809
                     /gene="T12H17.60"
                     /number=5
     intron          11810. .11897
                     /gene="T12H17.60"
                     /number=5
     exon            11898. .12053
                     /gene="T12H17.60"
                     /number=6

  Query Match             19.1%;  Score 186.4;  DB 8;  Length 89350;
  Best Local Similarity   63.0%;  Pred. No. 1.9e-41;
  Matches  305;  Conservative    0;  Mismatches  176;  Indels    3;  Gaps    1;

Qy     160 acgtcgtccacgtggtcgtcctgcaggttccaaaaacaaacccaaacctccgacgattat 219
           | |  | ||| | ||  | || || || ||||| |||||||| ||||| ||||  || ||
Db   77781 AAGAAGACCAAGAGGAAGACCAGCGGGATCCAAGAACAAACCAAAACCACCGATTATCAT 77840

Qy     220 aactagagatagtcctaacgtccttagatcacacgttcttgaagtcacctccggttcgga 279
             ||  | || ||  | ||||  |||||| | |||||  | ||  ||      || |  ||
Db   77841 CACACGGGACAGCGCAAACGCGCTTAGAACCCACGTGATGGAGATCGGAGATGGCTGCGA 77900

Qy     280 catatccgaggcagtctccacctacgccactcgtcgcggctgcggcgtttgcattataag 339
           | ||  |||    ||  |||| |  ||     | |||    |||||||||||| |||| ||
Db   77901 CTTAGTCGAAAGCGTTGCCACTTTTGCACGAAGACGCCAACGCGGCGTTTGCGTTATGAG 77960

Qy     340 cggcacgggtgcggtcactaacgtcacgatacggcaacctg---cggctccggctggtgg 396
           ||| || ||    || |||||||||||| ||||| || ||||   |   |||  ||  |||
Db   77961 CGGTACTGGAAATGTTACTAACGTCACTATACGTCAGCCTGGATCTCATCCTTCTCCTGG 78020

Qy     397 aggtgtgattaccctgcatggtcggtttgacattttgtctttgaccggtactgcgcttcc 456
             ||  |||  || || ||  |||| || ||| | ||| |  | ||  ||   || ||
Db   78021 CTCGGTAGTTAGTCTTCACGGAAGGTTCGAGATTCTATCTCTCTCAGGATCTTTTCTCCC 78080

Qy     457 accgcctgcaccaccgggagcaggaggtttgacggtgtatctagccggaggtcaaggaca 516
           || || || || ||  |||   || ||||  || || || || ||||| ||||||||
Db   78081 TCCTCCGGCTCCTCCTACAGCCACCGGATTGAGTGTTTACCTCGCTGGAGGACAAGGACA 78140

Qy     517 agttgtaggagggaatgtggctggttcgttaattgcttcgggaccggtagtgttgatggc 576
           || || ||||| |  || | |||| |||| |   | | || || || ||  | |||||
Db   78141 GGTGGTTGGAGGAAGCGTAGTTGGTCCGTTGTTATGTGCTGGTCCTGTCGTTGTCATGGC 78200

Qy     577 tgcttcttttgcaaacgcagtttatgataggttaccgattgaagaggaagaaaccccacc 636
           ||| ||||||   || || |  || || ||||| ||  | || || || || |  |   |
Db   78201 TGCGTCTTTTAGCAATGCGGCGTACGAAAGGTTGCCTTTAGAGGAAGATGAGATGCAGAC 78260

Qy     637 gccg 640
           ||||
Db   78261 GCCG 78264


Search completed: July 10, 2002, 21:20:06
Job time: 8294 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2002, 20:07:47 ; Search time 231.22 Seconds
(without alignments)
7232.396 Million cell updates/sec

Title: US-09-823-676-1
Perfect score: 974
Sequence: 1 ccccccgacctgcctctaca.....cgaaatgcgaatattaggtt 974

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	974	100.0	974	22	AAD05839	Arabidopsis thalia
2	974	100.0	974	22	AAF80408	Nucleotide sequenc
3	196.8	20.2	1020	22	AAD06672	A. thaliana transc
4	191.6	19.7	1307	21	AAC48248	Arabidopsis thalia
5	191.6	19.7	1310	21	AAC46844	Arabidopsis thalia
6	190.4	19.5	1097	21	AAC42265	Arabidopsis thalia
7	179.2	18.4	1192	21	AAC39286	Arabidopsis thalia
8	178.8	18.4	1190	21	AAC51552	Arabidopsis thalia
9	178.8	18.4	1251	24	AAS96327	Arabidopsis cDNA e
10	177.2	18.2	1303	21	AAC35116	Arabidopsis thalia
c 11	127.6	13.1	12319	21	AAZ89379	A. thaliana ACW1 g
12	42	4.3	1219	21	AAC44421	Arabidopsis thalia
13	42	4.3	1263	22	AAD05779	Arabidopsis thalia
14	42	4.3	1387	21	AAC38928	Arabidopsis thalia
15	42	4.3	1606	22	AAD06652	A. thaliana transc
16	38.6	4.0	1745	22	AAI29497	C883P determined c
17	37.4	3.8	1361	22	AAD02889	Dimorphotheca sinu
18	37	3.8	7696	22	AAS46399	Tumour suppressor
c 19	36	3.7	3062	23	ABL23912	Drosophila melanog
c 20	35.8	3.7	510	23	ABL22333	Drosophila melanog
21	35.8	3.7	1407	22	AAD06671	A. thaliana transc
c 22	35.8	3.7	2606	23	ABL22332	Drosophila melanog
23	35.8	3.7	4280	18	AAV74392	Staphylococcus aur
24	35.6	3.7	576	22	AAS01677	Human apolipoprote
c 25	35.2	3.6	5193	23	ABL23107	Drosophila melanog
26	35.2	3.6	9577	23	ABL23106	Drosophila melanog
c 27	34.8	3.6	330	24	AAS97212	Neisseria meningit
c 28	34.8	3.6	330	24	AAS97230	Neisseria meningit
c 29	34.8	3.6	1039	18	AAT83800	DNA encoding a gly
c 30	34.8	3.6	5276	22	AAD08671	Drosophila sp. mus
c 31	34.6	3.6	7320	23	ABL24159	Drosophila melanog
32	34.6	3.6	9718	23	ABL24158	Drosophila melanog
33	34.6	3.6	16831	22	AAK73455	Human immune/haema
c 34	34.4	3.5	4409	23	ABL17895	Drosophila melanog
c 35	34.4	3.5	5021	23	ABL11783	Drosophila melanog
c 36	34.4	3.5	7394	23	ABL17894	Drosophila melanog
c 37	34.4	3.5	7614	23	ABL11782	Drosophila melanog
38	34.2	3.5	781	23	ABL07225	Drosophila melanog
39	34.2	3.5	1075	22	ABA48937	Human breast cell
40	34.2	3.5	1075	22	ABA66852	Human foetal liver
41	34.2	3.5	1075	22	ABA33921	Probe #12387 for g
42	34.2	3.5	1075	22	AAK15288	Human brain expres
43	34.2	3.5	1075	22	AAK41011	Human bone marrow
44	34.2	3.5	1075	22	AAI21781	Probe #11714 for g
45	34.2	3.5	1075	22	AAI47064	Probe #15750 used

ALIGNMENTS

RESULT 1
AAD05839
ID AAD05839 standard; cDNA; 974 BP.
XX
AC AAD05839;
XX
DT 31-JUL-2001 (first entry)
XX
DE Arabidopsis thaliana transcription factor G1073 cDNA.
XX
KW Transcription factor; trait modification; structural characteristic;
KW developmental characteristic; gene therapy; agricultural biotechnology;
KW ss.
XX
OS Arabidopsis thaliana.
XX
FH Key Location/Qualifiers
FT CDS 62..874
FT /*tag= a
FT /product= "Transcription factor G1073"
XX
PN WO200136444-A1.
XX
PD 25-MAY-2001.
XX
PF 14-NOV-2000; 2000WO-US31325.
XX
PR 17-NOV-1999; 99US-0166228.
PR 17-APR-2000; 2000US-0197899.
PR 22-AUG-2000; 2000US-0227439.
XX
PA (MEND-) MENDEL BIOTECHNOLOGY INC.

PA (RIEC/) RIECHMANN J L.
PA (REUB/) REUBER L.
PA (KEDD/) KEDDIE J.
PA (RATC/) RATCLIFFE O.
PA (HEAR/) HEARD J.
PA (SAMA/) SAMAHA R.
PA (YUGG/) YU G.
PA (JIAN/) JIANG C.
XX
PI Riechmann JL, Reuber L, Keddie J, Ratcliffe O, Heard J, Samaha R;
PI Yu G, Jiang C;
XX
DR WPI; 2001-335996/35.
DR P-PSDB; AAE01961.
XX
PT Nucleic acids encoding plant transcription factor polypeptides, useful
PT for altering the developmental and structural characteristics of
PT plants, e.g. corn, potato and cotton plants -
XX
PS Claim 4; Page 102-103; 140pp; English.
XX
CC The present sequence is Arabidopsis thaliana transcription factor G1073
CC cDNA. The transcription factors are used to modify traits associated
CC with structural or developmental characteristics of plants, e.g.soybean,
CC wheat, corn, potato, cotton, rice, oilseed rape, sunflower, alfalfa,
CC sugar cane, turf, banana, blackberry, blueberry, strawberry, raspberry,
CC cantaloupe, carrot, cauliflower, coffee, cucumber, eggplant, grapes,
CC honey dew, lettuce, mango, melon, papaya, peas, watermelon, pineapple,
CC spinach, squash, sweet corn, tobacco, tomato, peppers, onion, rosaceous
CC fruits and/or vegetable brassicas when their expression levels are
CC altered. The manipulation of transcription factor levels in plants
CC offers great potential in agricultural biotechnology for modifying
CC plant's traits. The transcription factors are also used in gene therapy.
XX
SQ Sequence 974 BP; 222 A; 246 C; 279 G; 227 T; 0 other;

Query Match 100.0%; Score 974; DB 22; Length 974;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 974; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 ccccccgacctgcctctacagagacctgaagattccagaaccccacctgatcaaaaataa 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ccccccgacctgcctctacagagacctgaagattccagaaccccacctgatcaaaaataa 60

Qy      61 catggaacttaacagatctgaagcagacgaagcaaaggccgagaccactcccaccggtgg 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 catggaacttaacagatctgaagcagacgaagcaaaggccgagaccactcccaccggtgg 120

Qy     121 agccaccagctcagccacagcctctggctcttcctccggacgtcgtccacgtggtcgtcc 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 agccaccagctcagccacagcctctggctcttcctccggacgtcgtccacgtggtcgtcc 180

Qy     181 tgcaggttccaaaaacaaacccaaacctccgacgattataactagagatagtcctaacgt 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 tgcaggttccaaaaacaaacccaaacctccgacgattataactagagatagtcctaacgt 240

Qy     241 ccttagatcacacgttcttgaagtcacctccggttcggacatatccgaggcagtctccac 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 ccttagatcacacgttcttgaagtcacctccggttcggacatatccgaggcagtctccac 300

Qy     301 ctacgccactcgtcgcggctgcggcgtttgcattataagcggcacgggtgcggtcactaa 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 ctacgccactcgtcgcggctgcggcgtttgcattataagcggcacgggtgcggtcactaa 360

Qy     361 cgtcacgatacggcaacctgcggctccggctggtggaggtgtgattaccctgcatggtcg 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 cgtcacgatacggcaacctgcggctccggctggtggaggtgtgattaccctgcatggtcg 420

Qy     421 gtttgacattttgtctttgaccggtactgcgcttccaccgcctgcaccaccgggagcagg 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 gtttgacattttgtctttgaccggtactgcgcttccaccgcctgcaccaccgggagcagg 480

Qy     481 aggtttgacggtgtatctagccggaggtcaaggacaagttgtaggagggaatgtggctgg 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 aggtttgacggtgtatctagccggaggtcaaggacaagttgtaggagggaatgtggctgg 540

Qy     541 ttcgttaattgcttcgggaccggtagtgttgatggctgcttcttttgcaaacgcagttta 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 ttcgttaattgcttcgggaccggtagtgttgatggctgcttcttttgcaaacgcagttta 600

Qy     601 tgataggttaccgattgaagaggaagaaaccccaccgccgagaaccaccggggtgcagca 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 tgataggttaccgattgaagaggaagaaaccccaccgccgagaaccaccggggtgcagca 660

Qy     661 gcagcagccggaggcgtctcagtcgtcggaggttacggggagtggggcccaggcgtgtga 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 gcagcagccggaggcgtctcagtcgtcggaggttacggggagtggggcccaggcgtgtga 720

Qy     721 gtcaaacctccaaggtggaaatggtggaggaggtgttgctttctacaatcttggaatgaa 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 gtcaaacctccaaggtggaaatggtggaggaggtgttgctttctacaatcttggaatgaa 780

Qy     781 tatgaacaattttcaattctccgggggagatatttacggtatgagcggcggtagcggagg 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 tatgaacaattttcaattctccgggggagatatttacggtatgagcggcggtagcggagg 840

Qy     841 aggtggtggcggtgcgactagacccgcgttttagagttttagcgttttggtgacaccttt 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 aggtggtggcggtgcgactagacccgcgttttagagttttagcgttttggtgacaccttt 900

Qy     901 tgttgcgtttgcgtgtttgacctcaaactactaggctactagctatagcggttgcgaaat 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 tgttgcgtttgcgtgtttgacctcaaactactaggctactagctatagcggttgcgaaat 960

Qy     961 gcgaatattaggtt 974
          ||||||||||||||
Db     961 gcgaatattaggtt 974
```

RESULT 2
AAF80408
ID AAF80408 standard; cDNA; 974 BP.
XX
AC AAF80408;
XX
DT 29-JUN-2001 (first entry)
XX
DE Nucleotide sequence of a plant transcription factor G1073.
XX
KW Transcription factor; flowering time; transgenic plant; vernalisation;
KW plant development; plant physiology; flowering; ss.
XX
OS Arabidopsis thaliana.
XX
FH Key Location/Qualifiers
FT CDS 62..874
FT /*tag= a
FT /product= "transcription factor"
XX
PN WO200126459-A2.
XX
PD 19-APR-2001.
XX
PF 11-OCT-2000; 2000WO-US28141.
XX
PR 12-OCT-1999; 99US-0159464.
PR 08-NOV-1999; 99US-0164132.
PR 17-NOV-1999; 99US-0166228.
PR 17-APR-2000; 2000US-0197899.
PR 22-AUG-2000; 2000US-0227439.
XX
PA (MEND-) MENDEL BIOTECHNOLOGY INC.
PA (RATC/) RATCLIFFE O.
PA (HEAR/) HEARD J.
PA (SAMA/) SAMAHA R.

PA (CREE/) CREELMAN R.
PA (KEDD/) KEDDIE J.
PA (JIAN/) JIANG C.
PA (REUB/) REUBER L.
PA (RIEC/) RIECHMANN J L.
XX
PI Ratcliffe O, Heard J, Samaha R, Creelman R, Keddie J, Jiang C;
PI Reuber L, Riechmann JL;
XX
DR WPI; 2001-266398/27.
DR P-PSDB; AAB67838.
XX
PT New transgenic plant comprises a recombinant polynucleotide encoding a
PT plant transcription factor polypeptide and has a modified flowering
PT time or vernalization requirement -
XX
PS Example 7; Page 93-94; 108pp; English.
XX
CC The present sequence encodes a plant transcription factor protein
CC which modifies the flowering time of a plant. The polynucleotide
CC sequence is used to produce transgenic plants which have a modified
CC flowering time or a modified vernalisation requirement. The
CC polynucleotides and polypeptides are useful for modifying plant
CC development, physiology or biochemistry such that the modified plants
CC have a trait advantage over wild type plants. In particular they are
CC useful for accelerating, delaying or preventing flowering. The
CC polynucleotides are also useful as nucleic acid probes and primers.
CC They may be used to identify proteins that can modify the activity of
CC the transcription factor.
XX
SQ Sequence 974 BP; 222 A; 246 C; 279 G; 227 T; 0 other;

Query Match 100.0%; Score 974; DB 22; Length 974;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 974; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 ccccccgacctgcctctacagagacctgaagattccagaaccccacctgatcaaaaataa 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ccccccgacctgcctctacagagacctgaagattccagaaccccacctgatcaaaaataa 60

Qy     61 catggaacttaacagatctgaagcagacgaagcaaaggccgagaccactcccaccggtgg 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 catggaacttaacagatctgaagcagacgaagcaaaggccgagaccactcccaccggtgg 120

Qy    121 agccaccagctcagccacagcctctggctcttcctccggacgtcgtccacgtggtcgtcc 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 agccaccagctcagccacagcctctggctcttcctccggacgtcgtccacgtggtcgtcc 180

Qy    181 tgcaggttccaaaaacaaacccaaacctccgacgattataactagagatagtcctaacgt 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 tgcaggttccaaaaacaaacccaaacctccgacgattataactagagatagtcctaacgt 240

Qy    241 ccttagatcacacgttcttgaagtcacctccggttcggacatatccgaggcagtctccac 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ccttagatcacacgttcttgaagtcacctccggttcggacatatccgaggcagtctccac 300

Qy    301 ctacgccactcgtcgcggctgcggcgtttgcattataagcggcacgggtgcggtcactaa 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 ctacgccactcgtcgcggctgcggcgtttgcattataagcggcacgggtgcggtcactaa 360

Qy    361 cgtcacgatacggcaacctgcggctccggctggtggaggtgtgattaccctgcatggtcg 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 cgtcacgatacggcaacctgcggctccggctggtggaggtgtgattaccctgcatggtcg 420

Qy    421 gtttgacattttgtctttgaccggtactgcgcttccaccgcctgcaccaccgggagcagg 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 gtttgacattttgtctttgaccggtactgcgcttccaccgcctgcaccaccgggagcagg 480

Qy    481 aggtttgacggtgtatctagccggaggtcaaggacaagttgtaggagggaatgtggctgg 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 aggtttgacggtgtatctagccggaggtcaaggacaagttgtaggagggaatgtggctgg 540

Qy    541 ttcgttaattgcttcgggaccggtagtgttgatggctgcttcttttgcaaacgcagttta 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 ttcgttaattgcttcgggaccggtagtgttgatggctgcttcttttgcaaacgcagttta 600

Qy    601 tgataggttaccgattgaagaggaagaaaccccaccgccgagaaccaccggggtgcagca 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 tgataggttaccgattgaagaggaagaaaccccaccgccgagaaccaccggggtgcagca 660

Qy    661 gcagcagccggaggcgtctcagtcgtcggaggttacggggagtggggcccaggcgtgtga 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 gcagcagccggaggcgtctcagtcgtcggaggttacggggagtggggcccaggcgtgtga 720

Qy    721 gtcaaacctccaaggtggaaatggtggaggaggtgttgctttctacaatcttggaatgaa 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 gtcaaacctccaaggtggaaatggtggaggaggtgttgctttctacaatcttggaatgaa 780

Qy    781 tatgaacaattttcaattctccgggggagatatttacggtatgagcggcggtagcggagg 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 tatgaacaattttcaattctccgggggagatatttacggtatgagcggcggtagcggagg 840

Qy    841 aggtggtggcggtgcgactagacccgcgttttagagttttagcgttttggtgacaccttt 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 aggtggtggcggtgcgactagacccgcgttttagagttttagcgttttggtgacaccttt 900

Qy    901 tgttgcgtttgcgtgtttgacctcaaactactaggctactagctatagcggttgcgaaat 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 tgttgcgtttgcgtgtttgacctcaaactactaggctactagctatagcggttgcgaaat 960

Qy    961 gcgaatattaggtt 974
          |||||||||||||
Db    961 gcgaatattaggtt 974
```

RESULT 3
AAD06672
ID AAD06672 standard; cDNA; 1020 BP.
XX
AC AAD06672;
XX
DT 10-AUG-2001 (first entry)
XX
DE A. thaliana transcription factor G1068 homolog, G2657 cDNA.
XX
KW Plant transcription factor; phenotype; sugar sensing characteristic;
KW transgenic plant; plant yield; growth; germination; photosynthesis;
KW glyoxylate metabolism; respiration; pathogen response; wounding response;
KW cell cycle regulation; pigmentation; flowering; senescence; physiology;
KW storage organ; metabolism; ss.
XX
OS Arabidopsis thaliana.
XX
FH Key Location/Qualifiers
FT CDS 1..1020
FT /*tag= a
FT /product= "Transcription factor homolog"
XX
PN WO200135725-A1.
XX
PD 25-MAY-2001.
XX
PF 14-NOV-2000; 2000WO-US31414.
XX
PR 17-NOV-1999; 99US-0166228.
PR 17-APR-2000; 2000US-0197899.
PR 22-AUG-2000; 2000US-0227439.
XX
PA (MEND-) MENDEL BIOTECHNOLOGY INC.
PA (JIAN/) JIANG C.
PA (HEAR/) HEARD J.
PA (PINE/) PINEDA O.
PA (PILG/) PILGRIM M.
PA (ADAM/) ADAM L.
PA (RIEC/) RIECHMANN J L.

PA (YUGG/) YU G.
PA (SAMA/) SAMAHA R.
XX
PI Jiang C, Heard J, Pineda O, Pilgrim M, Adam L, Riechmann JL;
PI Yu G, Samaha R;
XX
DR WPI; 2001-335977/35.
DR P-PSDB; AAE02571.
XX
PT Nucleic acids encoding plant transcription factor polypeptides, useful
PT for altering the sugar sensing characteristics of plants and increasing
PT yield, e.g. corn, potato and cotton plants -
XX
PS Claim 4; Page 143-144; 151pp; English.
XX
CC The patent relates to polynucleotides encoding 35 plant transcription
CC factors which may be used to modify phenotype associated with a plant's
CC sugar sensing characteristics and increasing yield when their expression
CC level is altered. Sugars are central regulatory molecules that control
CC aspects of physiology, metabolism and development. Therefore the cDNAs
CC and proteins of the invention are useful for modifying the growth and
CC germination rates of plants, photosynthesis, glyoxylate metabolism,
CC respiration, starch and sucrose synthesis and degradation, pathogen
CC response, wounding response, cell cycle regulation, pigmentation,
CC flowering and senescence of plants and for modifying sink-source
CC relationships in seeds, tubers, roots, and other storage organs leading
CC to an increase in yield. The transcription factor polynucleotides and
CC polypeptides may be used to alter the structure and developmental
CC characteristics of plants such as soybean, wheat, corn, potato, cotton,
CC rice, oilseed, sunflower, alfalfa, sugarcane, turf, banana, blackberry,
CC blueberry, strawberry, raspberry, cantaloupe, carrot, cauliflower,
CC coffee, cucumber, eggplant, grapes, honey dew, lettuce, mango, melon,
CC onion, papaya, peas, peppers, pineapple, spinach, squash, sweet corn,
CC tobacco, tomato, watermelon, rosaceous fruits and/or vegetable brassicas.
CC The present sequence is a homolog of Arabidopsis thaliana transcription
CC factor cDNA.
XX
SQ Sequence 1020 BP; 290 A; 241 C; 277 G; 212 T; 0 other;

Query Match 20.2%; Score 196.8; DB 22; Length 1020;
Best Local Similarity 64.6%; Pred. No. 4.6e-52;
Matches 310; Conservative 0; Mismatches 167; Indels 3; Gaps 1;

```
Qy    160 acgtcgtccacgtggtcgtcctgcaggttccaaaaacaaacccaaacctccgacgattat 219
          | |  | ||| | ||  | || ||||| ||||| |||||||| ||| |||| |  || ||
Db    351 aagaaggccaagaggaagaccagcaggatccaagaacaaacctaaagctccaataatcat 410

Qy    220 aactagagatagtcctaacgtccttagatcacacgttcttgaagtcacctccggttcgga 279
          ||| ||||| ||  | ||||  ||| || | |||||  | ||  |      ||| |  ||
Db    411 aacaagagacagcgcaaacgcgcttcgaactcacgtcatggagataggagacggatgtga 470

Qy    280 catatccgaggcagtctccacctacgccactcgtcgcggctgcggcgtttgcattataag 339
          ||||   ||     |  | || | ||| |  || |||    | |||||||| |||| ||
Db    471 catagttgactgtatggctacgttcgctagacgccgccaaagaggcgtttgcgttatgag 530

Qy    340 cggcacgggtgcggtcactaacgtcacgatacggcaacctgcggctccggctggtggagg 399
          ||| || ||    || |||||||||| ||||| || ||||   | ||  |   |||
Db    531 cggtacaggaagcgttactaacgtcactatacgtcagcctggatcgccacc---tggctc 587

Qy    400 tgtgattaccctgcatggtcggtttgacattttgtctttgaccggtactgcgcttccacc 459
           ||| ||| ||| || || |||||||| ||  | ||| |  | ||  ||    | || ||
Db    588 ggtggttagccttcacggccggtttgaaatcctctctctttcgggatctttcttgcctcc 647

Qy    460 gcctgcaccaccgggagcaggaggtttgacggtgtatctagccggaggtcaaggacaagt 519
          |||||| || || | |||    ||  | |  || || |||||||||| ||||| || ||
Db    648 gcctgcgccgcctgcagccaccggactaagcgtttacctagccggaggacaagggcaggt 707

Qy    520 tgtaggagggaatgtggctggttcgttaattgcttcgggaccggtagtgttgatggctgc 579
           || ||||| | |||||  ||  | ||  |   |||||| || || ||| | ||||| ||
Db    708 cgttggaggtagtgtggtgggacctttgttgtgttcgggtcctgtggtggttatggcggc 767

Qy    580 ttcttttgcaaacgcagtttatgataggttaccgattgaagaggaagaaaccccaccgcc 639
```

```
          |||||||   || || |  || || ||| | ||  | ||||| || || |  |  ||||
Db    768 ttcttttagcaatgcggcgtacgaaaggctgcctttggaagaagatgagatgcagacgcc 827
```

RESULT 4
AAC48248
ID AAC48248 standard; DNA; 1307 BP.
XX
AC AAC48248;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 56797.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 19.7%; Score 191.6; DB 21; Length 1307;
Best Local Similarity 64.7%; Pred. No. 2.4e-50;
Matches 303; Conservative 0; Mismatches 159; Indels 6; Gaps 1;

```
Qy    161 cgtcgtccacgtggtcgtcctgcaggttccaaaaacaaacccaaacctccgacgattata 220
          ||| | ||| | ||| | ||  |||| ||||||||||||||||||| |||| |   || |
Db    357 cgtagaccaagaggtagaccaccaggatccaaaaacaaacccaaagctccaatctttgtg 416

Qy    221 actagagatagtcctaacgtccttagatcacacgttcttgaagtcacctccggttcggac 280
          || ||||| || || ||||  ||  |    || ||  | ||  || ||  |||    |||
Db    417 acaagagacagccccaacgcactccgtagccatgtcttggagatctccgacggcagtgac 476
```

```
Qy    281 atatccgaggcagtctccacctacgccactcgtcgcggctgcggcgtttgcattataagc 340
          |  ||||  || || |   || | | |  ||  |     |||||||||| || | |||
Db    477 gtcgccgacacaatcgctcacttctcaagacgcaggcaacgcggcgtttgcgttctcagc 536

Qy    341 ggcacgggtgcggtcactaacgtcacgatacggcaacctgcggctccggctggtggaggt 400
          || || ||  | ||| ||||||||||  | || ||| | || || ||       ||||||
Db    537 gggacaggctcagtcgctaacgtcaccctccgccaagccgccgcacc------aggaggt 590

Qy    401 gtgattaccctgcatggtcggtttgacattttgtctttgaccggtactgcgcttccaccg 460
          ||| |  | || || ||  ||||||| || || ||||| ||||| ||  || ||
Db    591 gtggtctctctccaaggcaggtttgaaatcttatctttaaccggtgctttcctccctgga 650

Qy    461 cctgcaccaccgggagcaggaggtttgacggtgtatctagccggaggtcaaggacaagtt 520
          ||| | ||||| ||  ||   ||||| ||||| ||  |||||| |  || || || ||
Db    651 ccttccccacccgggtcaaccggtttaacggtttacttagccggggtccagggtcaggtc 710

Qy    521 gtaggagggaatgtggctggttcgttaattgcttcgggaccggtagtgttgatggctgct 580
          || |||||  |  || |  ||  |  |  | ||    ||  ||||  || |||| ||||||
Db    711 gttggaggtagcgttgtaggcccactcttagccatagggtcggtcatggtgattgctgct 770

Qy    581 tcttttgcaaacgcagtttatgataggttaccgattgaagaggaagaa 628
          ||||  | |||||   |||||| || || || || ||||| || |||
Db    771 actttctctaacgctacttatgagagattgcccatggaagaagaggaa 818
```

RESULT 5
AAC46844
ID AAC46844 standard; DNA; 1310 BP.
XX
AC AAC46844;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 51614.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.

```
PR   27-JUL-1999;   99US-0145918.
PR   27-JUL-1999;   99US-0145919.
PR   28-JUL-1999;   99US-0145951.
PR   02-AUG-1999;   99US-0146386.
PR   02-AUG-1999;   99US-0146388.
PR   02-AUG-1999;   99US-0146389.
PR   03-AUG-1999;   99US-0147038.
PR   04-AUG-1999;   99US-0147204.
PR   04-AUG-1999;   99US-0147302.
PR   05-AUG-1999;   99US-0147192.
PR   05-AUG-1999;   99US-0147260.
PR   06-AUG-1999;   99US-0147303.
PR   06-AUG-1999;   99US-0147416.
PR   09-AUG-1999;   99US-0147493.
PR   09-AUG-1999;   99US-0147935.
PR   10-AUG-1999;   99US-0148171.
PR   11-AUG-1999;   99US-0148319.
PR   12-AUG-1999;   99US-0148341.
PR   13-AUG-1999;   99US-0148565.
PR   13-AUG-1999;   99US-0148684.
PR   16-AUG-1999;   99US-0149368.
PR   17-AUG-1999;   99US-0149175.
PR   18-AUG-1999;   99US-0149426.
PR   20-AUG-1999;   99US-0149722.
PR   20-AUG-1999;   99US-0149723.
PR   20-AUG-1999;   99US-0149929.
PR   23-AUG-1999;   99US-0149902.
PR   23-AUG-1999;   99US-0149930.
PR   25-AUG-1999;   99US-0150566.
PR   26-AUG-1999;   99US-0150884.
PR   27-AUG-1999;   99US-0151065.
PR   27-AUG-1999;   99US-0151066.
PR   27-AUG-1999;   99US-0151080.
PR   30-AUG-1999;   99US-0151303.
PR   31-AUG-1999;   99US-0151438.
PR   01-SEP-1999;   99US-0151930.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             19.7%;  Score 191.6;  DB 21;  Length 1310;
  Best Local Similarity   64.7%;  Pred. No. 2.4e-50;
  Matches  303;  Conservative    0;  Mismatches  159;  Indels    6;  Gaps    1;

Qy     161 cgtcgtccacgtggtcgtcctgcaggttccaaaaacaaacccaaacctccgacgattata 220
           ||| | ||| | ||| | ||  |||| ||||||||||||||||||| |||| |   || |
Db     357 cgtagaccaagaggtagaccaccaggatccaaaaacaaacccaaagctccaatctttgtg 416

Qy     221 actagagatagtcctaacgtccttagatcacacgttcttgaagtcacctccggttcggac 280
           || |||| || || ||||  ||  |    || ||  | ||  || ||  |||    |||
Db     417 acaagagacagccccaacgcactccgtagccatgtcttggagatctccgacggcagtgac 476

Qy     281 atatccgaggcagtctccacctacgccactcgtcgcggctgcggcgtttgcattataagc 340
           |  ||||  || || |   || | | |  ||  |     ||||||||||| || | |||
Db     477 gtcgccgacacaatcgctcacttctcaagacgcaggcaacgcggcgtttgcgttctcagc 536

Qy     341 ggcacgggtgcggtcactaacgtcacgatacggcaacctgcggctccggctggtggaggt 400
           || || ||  | ||| ||||||||||  | || ||| | || || ||      ||||||
Db     537 gggacaggctcagtcgctaacgtcaccctccgccaagccgccgcacc------aggaggt 590

Qy     401 gtgattaccctgcatggtcggtttgacattttgtctttgaccggtactgcgcttccaccg 460
           ||| |  | || || ||  ||||||| || || ||||| |||||| ||   || ||
Db     591 gtggtctctctccaaggcaggtttgaaatcttatctttaaccggtgctttcctccctgga 650

Qy     461 cctgcaccaccgggagcaggaggtttgacggtgtatctagccggaggtcaaggacaagtt 520
           ||| | ||||| ||  ||   ||||| ||||| ||  |||||||  |  || || || ||
Db     651 ccttccccacccgggtcaaccggtttaacggtttacttagccggggtccagggtcaggtc 710

Qy     521 gtaggagggaatgtggctggttcgttaattgcttcgggaccggtagtgttgatggctgct 580
           || |||||  |  || |  ||  |  | | ||    ||  ||||  || |||| ||||||
Db     711 gttggaggtagcgttgtaggcccactcttagccatagggtcggtcatggtgattgctgct 770

Qy     581 tcttttgcaaacgcagtttatgataggttaccgattgaagaggaagaa 628
           ||||  | |||||   |||||| || || || || ||||| || |||
Db     771 actttctctaacgctacttatgagagattgcccatggaagaagaggaa 818


RESULT   6
AAC42265
ID   AAC42265 standard; DNA; 1097 BP.
XX
AC   AAC42265;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 34898.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
```

PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.

```
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match            19.5%;  Score 190.4;  DB 21;  Length 1097;
  Best Local Similarity  64.5%;  Pred. No. 5.3e-50;
  Matches  303;  Conservative    0;  Mismatches  161;  Indels    6;  Gaps    1;

Qy    158 ggacgtcgtccacgtggtcgtcctgcaggttccaaaaacaaacccaaacctccgacgatt 217
          || ||||||||||||||  | ||  | || ||||| |||||||| |||||||||   |||
Db    459 ggtcgtcgtccacgtggcagaccaccgggatccaagaacaaaccgaaacctccggtaatt 518

Qy    218 ataactagagatagtcctaacgtccttagatcacacgttcttgaagtcacctccggttcg 277
          || ||  | || ||  | |||   || ||| | ||| ||||||||| ||   ||| |
Db    519 atcacgcgcgagagcgcaaacactctaagagctcacattcttgaagtaacaaacggctgc 578

Qy    278 gacatatccgaggcagtctccacctacgccactcgtcgcggctgcggcgtttgcattata 337
          ||  | | |||    ||  | || || ||   ||| |     | ||  | ||| || |
Db    579 gatgttttcgactgcgttgcgacttatgctcgtcggagacagcgagggatctgcgttctg 638

Qy    338 agcggcacgggtgcggtcactaacgtcacgatacggcaacctgcggctccggctggtgga 397
          ||||| |  ||  ||||||| |||||||  ||||| ||      | |  | || | ||||
Db    639 agcggtagcggaacggtcacgaacgtcagcatacgtca------gccatctgcggctgga 692

Qy    398 ggtgtgattaccctgcatggtcggtttgacattttgtctttgaccggtactgcgcttcca 457
          |  ||  | || || || ||   ||| || ||| | ||| |  ||||  |    |||||
Db    693 gcggttgtgacgctacaaggaacgttcgagattctttctctctccggatcgtttcttcct 752

Qy    458 ccgcctgcaccaccgggagcaggaggtttgacggtgtatctagccggaggtcaaggacaa 517
          || || ||||| || ||||||    |||||||  | |   |||||||||| ||||| ||
Db    753 cctccggcacctcccggagcaacgagtttgacaattttcttagccggaggacaaggtcag 812

Qy    518 gttgtaggagggaatgtggctggttcgttaattgcttcgggaccggtagtgttgatggct 577
          || || |||||  |  || |  |||  | | |  ||  | ||||||||  |  |||| ||
Db    813 gtggttggaggaagcgttgtgggtgagcttacggcggctggaccggtgattgtgattgca 872

Qy    578 gcttcttttgcaaacgcagtttatgataggttaccgattgaagaggaaga 627
          ||||| ||| | || |  | |||||| ||  | ||  | ||||| || ||
Db    873 gcttcgtttactaatgttgcttatgagagacttcctttagaagaagatga 922


RESULT   7
AAC39286
ID   AAC39286 standard; DNA; 1192 BP.
XX
AC   AAC39286;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 24070.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
```

PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 18.4%; Score 179.2; DB 21; Length 1192;
Best Local Similarity 58.4%; Pred. No. 2.1e-46;
Matches 354; Conservative 0; Mismatches 243; Indels 9; Gaps 2;

```
Qy      38 gaaccccacctgatcaaaaataacatggaacttaacagatctgaagcagacgaagcaaag 97
           || |   || | | ||  || ||||  | |      |   || |  |  |      | |
Db     158 gaccatgacttcaccaccaacaacagtggaagccctaatactcagactcagagccaagaa 217

Qy      98 gccgagaccactcccaccggtggagccaccagctcagccacagcctctggctcttcctcc 157
           |   ||| ||       ||    | |  |      | ||  | ||     |   ||  |
Db     218 gaacagaacagcagagacgagcaaccagctgttgaacccggatccggatccgggtctacg 277

Qy     158 ggacgtcgtccacgtggtcgtcctgcaggttccaaaaacaaacccaaacctccgacgatt 217
           || ||||||||  | ||| | ||| | |||||||| |||||||| ||   |||    ||
Db     278 ggtcgtcgtcctagaggtagacctcctggttccaagaacaaaccaaagagtccagttgtt 337

Qy     218 ataactagagatagtcctaacgtccttagatcacacgttcttgaagtcacctccggttcg 277
           | || | ||| || ||||||   ||       || ||||||||  |  |  | ||  |
Db     338 gttaccaaagaaagccctaactctctccagagccatgttcttgagattgctacgggagct 397

Qy     278 gacatatccgaggcagtctccacctacgccactcgtcgcggctgcggcgtttgcattata 337
           ||| |  | ||     |   | |||  ||   | | |||||| | |||||||   |  |
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2002, 22:14:33 ; Search time 56.06 Seconds
(without alignments)
462.792 Million cell updates/sec

Title: US-09-823-676-2
Perfect score: 1392
Sequence: 1 MELNRSEADEAKAETTPTGG.....IYGMSGGSGGGGGGATRPAF 270

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1392	100.0	270	2	T06118	hypothetical prote
2	706.5	50.8	311	2	F86341	hypothetical prote
3	646.5	46.4	292	2	C71448	hypothetical prote
4	645.5	46.4	302	2	H96792	unknown protein F1
5	629.5	45.2	317	2	D84890	probable AT-hook D
6	625.5	44.9	339	2	T06612	hypothetical prote
7	616	44.3	285	2	E84766	probable AT-hook D
8	605	43.5	324	2	T04576	hypothetical prote
9	581	41.7	265	2	T47898	hypothetical prote
10	572	41.1	310	2	T47695	hypothetical prote
11	554	39.8	257	2	B84860	hypothetical prote
12	417	30.0	206	2	G86279	F14L17.27 protein
13	274	19.7	296	2	G84747	AT-hook DNA-bindin
14	255.5	18.4	334	2	T04572	hypothetical prote
15	246.5	17.7	347	2	T06584	probable DNA-bindi
16	243	17.5	439	2	T52291	probable DNA-bindi
17	239.5	17.2	404	2	T05553	hypothetical prote
18	239	17.2	455	2	T05081	hypothetical prote
19	227	16.3	365	2	T06615	hypothetical prote
20	225	16.2	348	2	T02462	probable AT-hook D
21	218	15.7	345	2	T01348	hypothetical prote
22	211.5	15.2	348	2	T47923	probable DNA-bindi
23	162.5	11.7	574	2	A84782	hypothetical prote
24	146	10.5	405	2	T29167	hypothetical prote
25	139	10.0	307	2	T27609	hypothetical prote
26	137	9.8	694	2	F70868	hypothetical glyci
27	135.5	9.7	290	2	T23416	hypothetical prote
28	134.5	9.7	839	2	F75518	hypothetical prote
29	131	9.4	388	2	T29173	hypothetical prote
30	131	9.4	1381	2	E70806	hypothetical glyci
31	129.5	9.3	575	2	S35327	protein kinase sgg
32	129.5	9.3	1844	2	T51890	related to Nup98-N
33	129.5	9.3	13288	2	T03099	mucin, submaxillar
34	128.5	9.2	1585	2	T31611	hypothetical prote
35	128	9.2	2468	2	A83412	hypothetical prote
36	125.5	9.0	2038	2	A43742	female sterile hom
37	125	9.0	528	2	D46449	hypothetical prote
38	124.5	8.9	867	2	S57795	probable deoxyribo
39	124	8.9	1655	2	T13998	gene mastermind pr
40	124	8.9	3190	2	T13828	CREB-binding prote
41	122	8.8	615	2	H70589	hypothetical glyci
42	122	8.8	979	2	A35913	regulatory factor
43	121.5	8.7	670	2	T49510	fibroin-3 related
44	121.5	8.7	1231	2	S30185	insulin receptor s
45	121	8.7	418	2	B64924	hypothetical prote

ALIGNMENTS

RESULT 1
T06118
hypothetical protein F23E12.50 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 20-Jun-2000
C;Accession: T06118
R;Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Hoheis
submitted to the Protein Sequence Database, April 1999
A;Reference number: Z15485
A;Accession: T06118
A;Molecule type: DNA
A;Residues: 1-270 <BEV>
A;Cross-references: EMBL:AL022604; GSPDB:GN00062; ATSP:F23E12.50
A;Experimental source: cultivar Columbia; BAC clone F23E12
C;Genetics:
A;Gene: ATSP:F23E12.50
A;Map position: 4
C;Superfamily: Arabidopsis thaliana hypothetical protein T12H17.200

Query Match 100.0%; Score 1392; DB 2; Length 270;
Best Local Similarity 100.0%; Pred. No. 6.6e-85;
Matches 270; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MELNRSEADEAKAETTPTGGATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDSPNV 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MELNRSEADEAKAETTPTGGATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDSPNV 60

Qy      61 LRSHVLEVTSGSDISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGR 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 LRSHVLEVTSGSDISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGR 120

Qy     121 FDILSLTGTALPPPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVY 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 FDILSLTGTALPPPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVY 180

Qy     181 DRLPIEEEETPPPRTTGVQQQQPEASQSSEVTGSGAQACESNLQGGNGGGGVAFYNLGMN 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 DRLPIEEEETPPPRTTGVQQQQPEASQSSEVTGSGAQACESNLQGGNGGGGVAFYNLGMN 240

Qy     241 MNNFQFSGGDIYGMSGGSGGGGGGATRPAF 270
          ||||||||||||||||||||||||||||||
Db     241 MNNFQFSGGDIYGMSGGSGGGGGGATRPAF 270
```

RESULT 2
F86341
hypothetical protein F9H16.12 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001
C;Accession: F86341
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon

```
            schuelle@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
COMMENT     Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
            and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.
FEATURES            Location/Qualifiers
     source         1. .86710
                    /organism="Arabidopsis thaliana"
                    /variety="Columbia"
                    /db_xref="taxon:3702"
                    /chromosome="4"
     gene           173. .1890
                    /gene="F23E12.10"
     exon           173. .374
                    /gene="F23E12.10"
                    /number=1
     CDS            join(<173. .374,456. .919,1089. .1610,1678. .1890)
                    /gene="F23E12.10"
                    /codon_start=1
                    /product="hypothetical protein (fragment)"
                    /protein_id="CAA18726.1"
                    /db_xref="GI:3080407"
                    /db_xref="SPTREMBL:O65486"
                    /translation="FNAAVAGCFFAVESVLWPSSSTDSSTSLPNTTSMVILSAVTASV
                    VSEIGLGSEPAFKVPDYDFRSPGELPLYLLLGALCGLVSLALSRCTSSMTSAVDSLNK
                    DAGIPKAVFPVMGGLSVGIIALVYPEVLYWGFQNVDILLEKRPFVKGLSADLLLQLVA
                    VKIAATAWCRASGLVGGYYAPSLFIGGAAGMAYGKFIGLALAQNPDFNLSILEVASPQ
                    AYGLVGMAATLAGVCQVPLTAVLLLFELTQDYRIVLPLLGAVGMSSWITSGQSKRQET
                    RETKETRKRKSQEAVQSLTSSDDESSTNNLCEVESSLCLDDSLNQSEELPKSIFVSEA
                    MRTRFATVMMSTSLEEALTRMLIEKQSCALIVDPDNIFLGILTLSDIQEFSKARKEGN
                    NRPKDIFVNDICSRSGGKCKVPWTVTPDMDLLAAQTIMNKHELSHVAVVSGSIDAPRI
                    HPVGVLDRECITLTRR"
     intron         375. .455
                    /gene="F23E12.10"
                    /number=1
     exon           456. .919
                    /gene="F23E12.10"
                    /number=2
     intron         920. .1088
                    /gene="F23E12.10"
                    /number=2
     exon           1089. .1610
                    /gene="F23E12.10"
                    /number=3
     intron         1611. .1677
                    /gene="F23E12.10"
                    /number=3
     exon           1678. .1890
                    /gene="F23E12.10"
                    /number=4
     exon           3124. .3241
                    /gene="F23E12.20"
                    /number=1
     gene           3124. .4428
                    /gene="F23E12.20"
     CDS            join(3124. .3241,3323. .3465,3595. .3696,3945. .4107,
                    4211. .4428)
                    /gene="F23E12.20"
                    /note="similarity to dihydroflavonol-4-reductase, Gerbera
                    x sp., PIR2:S35189"
                    /codon_start=1
                    /product="putative protein"
                    /protein_id="CAA18727.1"
                    /db_xref="GI:3080408"
                    /db_xref="SPTREMBL:O65487"
                    /translation="MDQAKGKVCVTGASGFLASWLVKRLLLEGYEVIGTVRDPGNEKK
                    LAHLWKLEGAKERLRLVKADLMEEGSFDNAIMGCQGVFHTASPEILRPAIEGTLNVLR
                    SCRKNPSLKRVVLTSSSSTVWYALSKTLAEQAAWKFSEENGIDLVTVLPSFLVGPSLP
                    PDLCSTASDVLGLLKGETEKFQWHGQMGYVHIDDVARTHIVVFEHEAAQGRYICSSNV
                    ISLEELVSFLSARYPSLPIPKRFDTDPLN"
     intron         3242. .3322
                    /gene="F23E12.20"
                    /number=1
     exon           3323. .3465
                    /gene="F23E12.20"
                    /number=2
     intron         3466. .3594
                    /gene="F23E12.20"
                    /number=2
     exon           3595. .3696
                    /gene="F23E12.20"
                    /number=3
     intron         3697. .3944
                    /gene="F23E12.20"
                    /number=3
     exon           3945. .4107
                    /gene="F23E12.20"
                    /number=4
     intron         4108. .4210
                    /gene="F23E12.20"
                    /number=4
     exon           4211. .4428
                    /gene="F23E12.20"
                    /number=5
     gene           complement(join(4939. .5013,5109. .5195,5307. .5342,
                    5451. .5651,5725. .5811,6161. .6163))
                    /gene="F23E12.30"
     gene           4939. .6163
                    /gene="F23E12.30"
     CDS            complement(join(4939. .5013,5109. .5195,5307. .5342,
                    5451. .5651,5725. .5811,6161. .6163))
                    /gene="F23E12.30"
                    /note="Contains Clathrin adaptor complexes small chain
                    signature [VVYKRYASLYF]
                    contains EST gb:T44219"
                    /codon_start=1
                    /product="clathrin assembly protein AP19 homolog"
                    /protein_id="CAA18728.1"
                    /db_xref="GI:3080409"
                    /db_xref="SPTREMBL:O23685"
                    /translation="MIHFVLLVSRQGKVRLTKWYSPYAQKERSKVIRELSGVILNRGP
                    KLCNFVEWRGYKVVYKRYASLYFCMCIDQEDNELEVLEIIHHYVEILDRYFGSVCELD
                    LIFNFHKAYYILDELLIAGELQESSKKTVARIISAQDQLVEVAKEEASSISNIIAQAT
                    NR"
     exon           complement(4939. .5013)
                    /gene="F23E12.30"
                    /number=1
     intron         complement(5014. .5108)
                    /gene="F23E12.30"
                    /number=1
     exon           complement(5109. .5195)
                    /gene="F23E12.30"
                    /number=2
     intron         complement(5196. .5306)
                    /gene="F23E12.30"
                    /number=2
     exon           complement(5307. .5342)
                    /gene="F23E12.30"
                    /number=3
     intron         complement(5343. .5450)
                    /gene="F23E12.30"
                    /number=3
     exon           complement(5451. .5651)
                    /gene="F23E12.30"
                    /number=4
     intron         complement(5652. .5724)
                    /gene="F23E12.30"
                    /number=4
     exon           complement(5725. .5811)
                    /gene="F23E12.30"
                    /number=5
     intron         complement(5812. .6160)
                    /gene="F23E12.30"
                    /number=5
```

```
     exon            complement(6161. .6163)
                     /gene="F23E12.30"
                     /number=6
     exon            complement(7352. .7582)
                     /gene="F23E12.40"
                     /number=1
     gene            7352. .7582
                     /gene="F23E12.40"
     gene            complement(7352. .7582)
                     /gene="F23E12.40"
     CDS             complement(7352. .7582)
                     /gene="F23E12.40"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAA18729.1"
                     /db_xref="GI:3080410"
                     /db_xref="SPTREMBL:O65488"
                     /translation="MPLVRKNGVVMRIGSPMLHESENLVGDGVEAPPRGERDEIGEEV
                     GGETNVVSESCGGEEGAQDAAKTTPEGEADRI"
     gene            8233. .9045
                     /gene="F23E12.50"
     gene            complement(8233. .9045)
                     /gene="F23E12.50"
     exon            complement(8233. .9045)
                     /gene="F23E12.50"
                     /number=1
     CDS             complement(8233. .9045)
                     /gene="F23E12.50"
                     /note="similarity to DNA-binding protein PD1, Pisum
                     sativum"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAA18730.1"

  Query Match             97.8%;  Score 952.8;  DB 8;  Length 86710;
  Best Local Similarity   99.7%;  Pred. No. 2.1e-258;
  Matches  965;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy       7 gacctgcctctacagagacctgaagattccagaaccccacctgatcaaaaataacatgga 66
           || |||| ||||||||||||||||||||||||||||||||||| ||||||||||||
Db    9099 GAACTGCATCTACAGAGACCTGAAGATTCCAGAACCCCACCTGATC-AAAATAACATGGA 9041

Qy      67 acttaacagatctgaagcagacgaagcaaaggccgagaccactcccaccggtggagccac 126
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9040 ACTTAACAGATCTGAAGCAGACGAAGCAAAGGCCGAGACCACTCCCACCGGTGGAGCCAC 8981

Qy     127 cagctcagccacagcctctggctcttcctccggacgtcgtccacgtggtcgtcctgcagg 186
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8980 CAGCTCAGCCACAGCCTCTGGCTCTTCCTCCGGACGTCGTCCACGTGGTCGTCCTGCAGG 8921

Qy     187 ttccaaaaacaaacccaaacctccgacgattataactagagatagtcctaacgtccttag 246
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8920 TTCCAAAAACAAACCCAAACCTCCGACGATTATAACTAGAGATAGTCCTAACGTCCTTAG 8861

Qy     247 atcacacgttcttgaagtcacctccggttcggacatatccgaggcagtctccacctacgc 306
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8860 ATCACACGTTCTTGAAGTCACCTCCGGTTCGGACATATCCGAGGCAGTCTCCACCTACGC 8801

Qy     307 cactcgtcgcggctgcggcgtttgcattataagcggcacgggtgcggtcactaacgtcac 366
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8800 CACTCGTCGCGGCTGCGGCGTTTGCATTATAAGCGGCACGGGTGCGGTCACTAACGTCAC 8741

Qy     367 gatacggcaacctgcggctccggctggtggaggtgtgattaccctgcatggtcggtttga 426
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8740 GATACGGCAACCTGCGGCTCCGGCTGGTGGAGGTGTGATTACCCTGCATGGTCGGTTTGA 8681

Qy     427 cattttgtctttgaccggtactgcgcttccaccgcctgcaccaccgggagcaggaggttt 486
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8680 CATTTTGTCTTTGACCGGTACTGCGCTTCCACCGCCTGCACCACCGGGAGCAGGAGGTTT 8621

Qy     487 gacggtgtatctagccggaggtcaaggacaagttgtaggagggaatgtggctggttcgtt 546
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8620 GACGGTGTATCTAGCCGGAGGTCAAGGACAAGTTGTAGGAGGGAATGTGGCTGGTTCGTT 8561

Qy     547 aattgcttcgggaccggtagtgttgatggctgcttcttttgcaaacgcagtttatgatag 606
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8560 AATTGCTTCGGGACCGGTAGTGTTGATGGCTGCTTCTTTTGCAAACGCAGTTTATGATAG 8501

Qy     607 gttaccgattgaagaggaagaaaccccaccgccgagaaccaccggggtgcagcagcagca 666
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8500 GTTACCGATTGAAGAGGAAGAAACCCCACCGCCGAGAACCACCGGGGTGCAGCAGCAGCA 8441

Qy     667 gccggaggcgtctcagtcgtcggaggttacggggagtggggcccaggcgtgtgagtcaaa 726
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8440 GCCGGAGGCGTCTCAGTCGTCGGAGGTTACGGGGAGTGGGGCCCAGGCGTGTGAGTCAAA 8381

Qy     727 cctccaaggtggaaatggtggaggaggtgttgctttctacaatcttggaatgaatatgaa 786
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8380 CCTCCAAGGTGGAAATGGTGGAGGAGGTGTTGCTTTCTACAATCTTGGAATGAATATGAA 8321

Qy     787 caattttcaattctccgggggagatatttacggtatgagcggcggtagcggaggaggtgg 846
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8320 CAATTTTCAATTCTCCGGGGGAGATATTTACGGTATGAGCGGCGGTAGCGGAGGAGGTGG 8261

Qy     847 tggcggtgcgactagacccgcgttttagagttttagcgttttggtgacacctttttgttgc 906
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8260 TGGCGGTGCGACTAGACCCGCGTTTTAGAGTTTTAGCGTTTTGGTGACACCTTTTGTTGC 8201

Qy     907 gtttgcgtgtttgacctcaaactactaggctactagctatagcggttgcgaaatgcgaat 966
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8200 GTTTGCGTGTTTGACCTCAAACTACTAGGCTACTAGCTATAGCGGTTGCGAAATGCGAAT 8141

Qy     967 attaggtt 974
           ||||||||
Db    8140 ATTAGGTT 8133

RESULT   2
ATCHRIV83
LOCUS       ATCHRIV83              197859 bp    DNA     linear   PLN 16-MAR-2000
DEFINITION  Arabidopsis thaliana DNA chromosome 4, contig fragment No. 83.
ACCESSION   AL161587
VERSION     AL161587.2  GI:7270470
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 16641 to 103350)
  AUTHORS   Hilbert,H., Braun,M., Holzer,E., Brandt,A., Duesterhoeft,A.,
            Mewes,H.W., Lemcke,K. and Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   2  (bases 98352 to 125759)
  AUTHORS   Robben,J., Grymonprez,B., Volckaert,G., Mewes,H.W., Lemcke,K. and
            Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   3  (bases 120761 to 197859)
  AUTHORS   Rose,M., Hempel,S., Entian,K.-D., Mewes,H.W., Lemcke,K. and
            Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   4  (bases 1 to 24256)
  AUTHORS   Murphy,G., Ridley,P., Hudson,S., Mewes,H.W., Lemcke,K. and
            Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   5  (bases 1 to 197859)
  AUTHORS   EU Arabidopsis sequencing,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
COMMENT     Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2002, 22:19:08 ; Search time 103.68 Seconds
(without alignments)
450.508 Million cell updates/sec

Title: US-09-823-676-2
Perfect score: 1392
Sequence: 1 MELNRSEADEAKAETTPTGG.....IYGMSGGSGGGGGGATRPAF 270

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1392	100.0	270	10	O65489	O65489 arabidopsis
2	706.5	50.8	311	10	Q9S7C9	Q9s7c9 arabidopsis
3	646.5	46.4	292	10	O23620	O23620 arabidopsis
4	645.5	46.4	302	10	Q9C9K7	Q9c9k7 arabidopsis
5	637.5	45.8	315	10	Q9SR17	Q9sr17 arabidopsis
6	629.5	45.2	317	10	O22130	O22130 arabidopsis
7	625.5	44.9	339	10	Q9SZ70	Q9sz70 arabidopsis
8	616	44.3	285	10	O82166	O82166 arabidopsis
9	605	43.5	324	10	O49662	O49662 arabidopsis
10	581	41.7	265	10	Q9LZX7	Q9lzx7 arabidopsis
11	572	41.1	310	10	Q9M2S3	Q9m2s3 arabidopsis
12	554	39.8	257	10	Q9SJG4	Q9sjg4 arabidopsis
13	499.5	35.9	276	10	Q9LTA2	Q9lta2 arabidopsis
14	417	30.0	206	10	Q9M9R4	Q9m9r4 arabidopsis
15	407	29.2	265	10	Q941Y9	Q941y9 oryza sativ
16	289.5	20.8	404	10	Q9LVB0	Q9lvb0 arabidopsis
17	289	20.8	771	10	Q9XHY4	Q9xhy4 oryza sativ
18	283	20.3	405	10	Q9AYM0	Q9aym0 oryza sativ
19	282	20.3	351	10	Q94F52	Q94f52 arabidopsis
20	274	19.7	296	10	O22812	O22812 arabidopsis
21	274	19.7	351	10	O23142	O23142 arabidopsis
22	264.5	19.0	419	10	Q9FHM5	Q9fhm5 arabidopsis
23	259	18.6	826	10	Q9SH33	Q9sh33 arabidopsis
24	258.5	18.6	300	10	Q9ZRR7	Q9zrr7 antirrhinum
25	255.5	18.4	334	10	O49658	O49658 arabidopsis
26	248.5	17.9	386	10	Q9FIR1	Q9fir1 arabidopsis
27	246.5	17.7	347	10	O04696	O04696 pisum sativ
28	243	17.5	439	10	O49350	O49350 arabidopsis
29	243	17.5	439	10	Q940I0	Q940i0 arabidopsis
30	242.5	17.4	334	10	O04695	O04695 pisum sativ
31	239.5	17.2	404	10	Q9SB31	Q9sb31 arabidopsis
32	239	17.2	455	10	O49694	O49694 arabidopsis
33	227	16.3	365	10	Q9SZ73	Q9sz73 arabidopsis
34	225	16.2	348	10	O80834	O80834 arabidopsis
35	221	15.9	418	10	Q9SR16	Q9sr16 arabidopsis
36	218	15.7	345	10	O81321	O81321 arabidopsis
37	211.5	15.2	348	10	Q9M2D3	Q9m2d3 arabidopsis
38	162.5	11.7	574	10	Q9SJQ2	Q9sjq2 arabidopsis
39	148	10.6	848	5	Q9V889	Q9v889 drosophila
40	146	10.5	405	5	Q23057	Q23057 caenorhabdi
41	143.5	10.3	818	6	Q9N1P0	Q9n1p0 bos taurus
42	139	10.0	1468	5	Q9GUB5	Q9gub5 galleria me
43	138.5	9.9	265	5	Q23347	Q23347 caenorhabdi
44	137	9.8	694	16	O53212	O53212 mycobacteri
45	136.5	9.8	1610	16	Q92KQ8	Q92kq8 rhizobium m

ALIGNMENTS

RESULT 1
O65489
ID O65489 PRELIMINARY; PRT; 270 AA.
AC O65489;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 26.8 KDA PROTEIN.
GN F23E12.50 OR AT4G35390.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA Hoheisel J., Jesse T., Heijnen L., Vos P., Mewes H.W., Mayer K.F.X.,
RA Schueller C.;
RL Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AL022604; CAA18730.1; -.
DR EMBL; AL161587; CAB80256.1; -.
KW Hypothetical protein.
SQ SEQUENCE 270 AA; 26774 MW; 4C21487C2719E114 CRC64;

Query Match 100.0%; Score 1392; DB 10; Length 270;
Best Local Similarity 100.0%; Pred. No. 4.7e-88;
Matches 270; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MELNRSEADEAKAETTPTGGATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDSPNV 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MELNRSEADEAKAETTPTGGATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDSPNV 60

Qy      61 LRSHVLEVTSGSDISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGR 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 LRSHVLEVTSGSDISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGR 120

Qy     121 FDILSLTGTALPPPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVY 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 FDILSLTGTALPPPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVY 180

Qy     181 DRLPIEEEETPPPRTTGVQQQQPEASQSSEVTGSGAQACESNLQGGNGGGGVAFYNLGMN 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 DRLPIEEEETPPPRTTGVQQQQPEASQSSEVTGSGAQACESNLQGGNGGGGVAFYNLGMN 240

Qy     241 MNNFQFSGGDIYGMSGGSGGGGGGATRPAF 270
           ||||||||||||||||||||||||||||||
Db     241 MNNFQFSGGDIYGMSGGSGGGGGGATRPAF 270


RESULT   2
Q9S7C9
ID   Q9S7C9      PRELIMINARY;      PRT;   311 AA.
AC   Q9S7C9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   ESCAROLA.
GN   ESC OR F9H16.12.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Weigel D., Ahn J.H., Blazquez M.A., Borevitz J., Christensen S.K.,
RA   Fankhauser C., Ferrandiz C., Kardailsky I., Neff M.M., Nguyen J.T.,
RA   Sato S., Wang Z., Xia Y., Dixon R.A., Harrison M.J., Lamb C.J.,
RA   Yanofsky M.F., Chory J.;
RT   "Activation Tagging in Arabidopsis.";
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,
RA   Altafi H., Araujo R., Huizar L., Rowley D., Buehler E., Dunn P.,
RA   Gonzalez A., Kremenetskaia I., Kim C., Lenz C., Li J., Liu S.,
RA   Luros S., Schwartz J., Shinn P., Toriumi M., Vysotskaia V.S.,
RA   Walker M., Yu G., Ecker J., Theologis A., Davis R.W.;
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF194974; AAF07197.1; -.
DR   EMBL; AC007369; AAD30602.1; -.
DR   InterPro; IPR000637; AT_hook.
DR   Pfam; PF02178; AT_hook; 1.
DR   SMART; SM00384; AT_hook; 1.
SQ   SEQUENCE   311 AA;  31842 MW;  A80B445C9776EB7D CRC64;


  Query Match             50.8%;  Score 706.5;  DB 10;  Length 311;
  Best Local Similarity   53.0%;  Pred. No. 5e-41;
  Matches  151;  Conservative   25;  Mismatches   56;  Indels   53;  Gaps    6;

Qy       7 EADEAKAETTPTGGATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDSPNVLRSHVL 66
           |:| :  :    |   |    | |: |:|||||| ||||| ||| |:|||||| ||||||
Db      59 ESDHSNKDHHQQGRPDSDPNTSSSAPGKRPRGRPPGSKNKAKPPIIVTRDSPNALRSHVL 118

Qy      67 EVTSGSDISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAP-------AGGGVITLHG 119
           ||: |:|| |:||||| ||| || :: | | |:|||:||| |        ||||:|||| 
Db     119 EVSPGADIVESVSTYARRRGRGVSVLGGNGTVSNVTLRQPVTPGNGGGVSGGGGVVTLHG 178

Qy     120 RFDILSLTGTALPPPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAV 179
           ||:||||||| |||||||||||:::|||||||||||:|   |||| ||:||||||:|||
Db     179 RFEILSLTGTVLPPPAPPGAGGLSIFLAGGQGQVVGGSVVAPLIASAPVILMAASFSNAV 238

Qy     180 YDRLPIEEEETP----------PPRTTGVQQQQPEASQSSEVTGSGAQACESNLQGGNG 228
           ::|||||||            ||:     || | ||  | ||| |         |||
Db     239 FERLPIEEEEEEGGGGGGGGGGPPQ----MQQAPSASPPSGVTGQGQ-------LGGNV 287

Qy     229 GGGVAFYNLGMNMNNFQFSGGDIYGMSGGS---GGGGGGATRPAF 270
           ||                      || ||     | | |  :|| |
Db     288 GG---------------------YGFSGDPHLLGWGAGTPSRPPF 311


RESULT   3
O23620
ID   O23620      PRELIMINARY;      PRT;   292 AA.
AC   O23620;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 29.7 KDA PROTEIN.
GN   AT4G17800.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98121113; PubMed=9461215;
RA   Bevan M., Bancroft I., Bent E., Love K., Goodman H., Dean C.,
RA   Bergkamp R., Dirkse W., van Staveren M., Stiekema W., Drost L.,
RA   Ridley P., Hudson S.A., Patel K., Murphy G., Piffanelli P., Wedler H.,
RA   Wedler E., Wambutt R., Weitzenegger T., Pohl T.M., Terryn N.,
RA   Gielen J., Villarroel R., De Clerck R., van Montagu M., Lecharny A.,
RA   Auborg S., Gy I., Kreis M., Lao N., Kavanagh T., Hempel S., Kotter P.,
RA   Entian K.D., Rieger M., Schaeffer M., Funk B., Mueller-Auer S.,
RA   Silvey M., James R., Montfort A., Pons A., Puigdomenech P., Douka A.,
RA   Voukelatou E., Milioni D., Hatzopoulos P., Piravandi E., Obermaier B.,
RA   Hilbert H., Duesterhoft A., Moores T., Jones J.D.G., Eneva T.,
RA   Palme K., Benes V., Rechman S., Ansorge W., Cooke R., Berger C.,
RA   Delseny M., Voet M., Volckaert G., Mewes H.W., Klosterman S.,
RA   Schueller C., Chalwatzis N.;
RT   "Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of
RT   Arabidopsis thaliana.";
RL   Nature 391:485-488(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z97344; CAB10560.1; -.
DR   EMBL; AL161547; CAB78783.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   292 AA;  29724 MW;  2A948700FA543C4F CRC64;


  Query Match             46.4%;  Score 646.5;  DB 10;  Length 292;
  Best Local Similarity   52.7%;  Pred. No. 5.9e-37;
  Matches  136;  Conservative   32;  Mismatches   47;  Indels   43;  Gaps    8;

Qy      19 GGATSSATASGSSS-----GRRPRGRPAGSKNKPKPPTIITRDSPNVLRSHVLEVTSGSD 73
           ||: ||    |        |||||||| |||||||| ||||:| | ||:|:||||:| |
Db      62 GGSGSSGGGGGHGGGGDVVGRRPRGRPPGSKNKPKPPVIITRESANTLRAHILEVTNGCD 121

Qy      74 ISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGRFDILSLTGTALPP 133
           : : |:||| ||  |:|::||:| ||||:||||:  | | |:|| | |:||||:|: |||
Db     122 VFDCVATYARRRQRGICVLSGSGTVTNVSIRQPS--AAGAVVTLQGTFEILSLSGSFLPP 179

Qy     134 PAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVYDRLPIEEEETPPP 193
           ||||||  ||::||||||||||||:| | | |:|||:::|||| |  |:|||:||:|
Db     180 PAPPGATSLTIFLAGGQGQVVGGSVVGELTAAGPVIVIAASFTNVAYERLPLEEDE---- 235

Qy     194 RTTGVQQQQPEASQSSEVTGSGAQACESNLQGGN--------GGGGVAFYNLGMNMN-NF 244
               ||||          | |:        |||        ||||: |:|| |||  |
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2002, 22:14:33 ; Search time 56.06 Seconds
(without alignments)
462.792 Million cell updates/sec

Title: US-09-823-676-2
Perfect score: 1392
Sequence: 1 MELNRSEADEAKAETTPTGG.....IYGMSGGSGGGGGGATRPAF 270

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1392	100.0	270	2	T06118	hypothetical prote
2	706.5	50.8	311	2	F86341	hypothetical prote
3	646.5	46.4	292	2	C71448	hypothetical prote
4	645.5	46.4	302	2	H96792	unknown protein F1
5	629.5	45.2	317	2	D84890	probable AT-hook D
6	625.5	44.9	339	2	T06612	hypothetical prote
7	616	44.3	285	2	E84766	probable AT-hook D
8	605	43.5	324	2	T04576	hypothetical prote
9	581	41.7	265	2	T47898	hypothetical prote
10	572	41.1	310	2	T47695	hypothetical prote
11	554	39.8	257	2	B84860	hypothetical prote
12	417	30.0	206	2	G86279	F14L17.27 protein
13	274	19.7	296	2	G84747	AT-hook DNA-bindin
14	255.5	18.4	334	2	T04572	hypothetical prote
15	246.5	17.7	347	2	T06584	probable DNA-bindi
16	243	17.5	439	2	T52291	probable DNA-bindi
17	239.5	17.2	404	2	T05553	hypothetical prote
18	239	17.2	455	2	T05081	hypothetical prote
19	227	16.3	365	2	T06615	hypothetical prote
20	225	16.2	348	2	T02462	probable AT-hook D
21	218	15.7	345	2	T01348	hypothetical prote
22	211.5	15.2	348	2	T47923	probable DNA-bindi
23	162.5	11.7	574	2	A84782	hypothetical prote
24	146	10.5	405	2	T29167	hypothetical prote
25	139	10.0	307	2	T27609	hypothetical prote
26	137	9.8	694	2	F70868	hypothetical glyci
27	135.5	9.7	290	2	T23416	hypothetical prote
28	134.5	9.7	839	2	F75518	hypothetical prote
29	131	9.4	388	2	T29173	hypothetical prote
30	131	9.4	1381	2	E70806	hypothetical glyci
31	129.5	9.3	575	2	S35327	protein kinase sgg
32	129.5	9.3	1844	2	T51890	related to Nup98-N
33	129.5	9.3	13288	2	T03099	mucin, submaxillar
34	128.5	9.2	1585	2	T31611	hypothetical prote
35	128	9.2	2468	2	A83412	hypothetical prote
36	125.5	9.0	2038	2	A43742	female sterile hom
37	125	9.0	528	2	D46449	hypothetical prote
38	124.5	8.9	867	2	S57795	probable deoxyribo
39	124	8.9	1655	2	T13998	gene mastermind pr
40	124	8.9	3190	2	T13828	CREB-binding prote
41	122	8.8	615	2	H70589	hypothetical glyci
42	122	8.8	979	2	A35913	regulatory factor
43	121.5	8.7	670	2	T49510	fibroin-3 related
44	121.5	8.7	1231	2	S30185	insulin receptor s
45	121	8.7	418	2	B64924	hypothetical prote

ALIGNMENTS

RESULT 1
T06118
hypothetical protein F23E12.50 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 20-Jun-2000
C;Accession: T06118
R;Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Hoheis
submitted to the Protein Sequence Database, April 1999
A;Reference number: Z15485
A;Accession: T06118
A;Molecule type: DNA
A;Residues: 1-270 <BEV>
A;Cross-references: EMBL:AL022604; GSPDB:GN00062; ATSP:F23E12.50
A;Experimental source: cultivar Columbia; BAC clone F23E12
C;Genetics:
A;Gene: ATSP:F23E12.50
A;Map position: 4
C;Superfamily: Arabidopsis thaliana hypothetical protein T12H17.200

Query Match 100.0%; Score 1392; DB 2; Length 270;
Best Local Similarity 100.0%; Pred. No. 6.6e-85;
Matches 270; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MELNRSEADEAKAETTPTGGATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDSPNV 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MELNRSEADEAKAETTPTGGATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDSPNV 60

Qy      61 LRSHVLEVTSGSDISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGR 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 LRSHVLEVTSGSDISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGR 120

Qy     121 FDILSLTGTALPPPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVY 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 FDILSLTGTALPPPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVY 180

Qy     181 DRLPIEEEETPPPRTTGVQQQQPEASQSSEVTGSGAQACESNLQGGNGGGGVAFYNLGMN 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 DRLPIEEEETPPPRTTGVQQQQPEASQSSEVTGSGAQACESNLQGGNGGGGVAFYNLGMN 240

Qy     241 MNNFQFSGGDIYGMSGGSGGGGGGATRPAF 270
          ||||||||||||||||||||||||||||||
Db     241 MNNFQFSGGDIYGMSGGSGGGGGGATRPAF 270
```

RESULT 2
F86341
hypothetical protein F9H16.12 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001
C;Accession: F86341
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon

```
Db    398 gacgtggcggaaagcttaaacgcctttgctcgtagacgcggccggggcgtttcggtgctg 457

Qy    338 agcggcacgggtgcggtcactaacgtcacgatacggcaacctgcggctccggctggtgga 397
          ||||| |  |||  ||| ||||| || ||  | || || ||||| ||  |        ||
Db    458 agcggtagtggtttggttactaatgttactctgcgtcagcctgctgcatc------cggt 511

Qy    398 ggtgtgattaccctgcatggtcggtttgacattttgtctttgaccggtactgcgcttcca 457
          || ||  |||   | | ||||| |||||| || |||||| ||   ||  ||   |||||
Db    512 ggagttgttagtttacgtggtcagtttgagatcttgtctatgtgtggggcttttcttcct 571

Qy    458 ccgcctg---caccaccgggagcaggaggtttgacggtgtatctagccggaggtcaagga 514
           || |||   | ||  | | ||| |  ||||| ||  | ||  |||| |||| ||||||
Db    572 acgtctggctctcctgctgcagccgctggtttaaccatttacttagctggagctcaaggt 631

Qy    515 caagttgtaggagggaatgtggctggttcgttaattgcttcgggaccggtagtgttgatg 574
          |||||||| |||||    || |||||  || | ||||| || ||||| ||  |  ||||
Db    632 caagttgtgggaggtggagttgctggcccgcttattgcctctggacccgttattgtgata 691

Qy    575 gctgcttctttttgcaaacgcagtttatgataggttaccgattgaagaggaagaaacccca 634
          |||||| | |||   || ||   |||||| |||||||||||||| || ||| ||   | |
Db    692 gctgctacgttttgcaatgccacttatgagaggttaccgattgaggaagaacaacagcaa 751

Qy    635 ccgccg 640
             || |
Db    752 gagcag 757
```

RESULT 8
AAC51552
ID AAC51552 standard; DNA; 1190 BP.
XX
AC AAC51552;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 68931.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.

PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 18.4%; Score 178.8; DB 21; Length 1190;
Best Local Similarity 63.2%; Pred. No. 2.8e-46;
Matches 311; Conservative 0; Mismatches 172; Indels 9; Gaps 2;

```
Qy    152 tcctccggacgtcgtccacgtggtcgtcctgcaggttccaaaaacaaacccaaacctccg 211
          ||  | || ||||||||  | ||| | ||| | |||||||| |||||||| ||   |||
Db    270 tctacgggtcgtcgtcctagaggtagacctcctggttccaagaacaaaccaaagagtcca 329

Qy    212 acgattataactagagatagtcctaacgtccttagatcacacgttcttgaagtcacctcc 271
              || | || | ||| || |||||||   ||       || ||||||||  |  |  |
Db    330 gttgttgttaccaaagaaagccctaactctctccagagccatgttcttgagattgctacg 389

Qy    272 ggttcggacatatccgaggcagtctccacctacgccactcgtcgcggctgcggcgtttgc 331
          ||  | ||| |  | ||     |   | |||  ||   | | |||||| | |||||||
Db    390 ggagctgacgtggcggaaagcttaaacgcctttgctcgtagacgcggccggggcgtttcg 449

Qy    332 attataagcggcacgggtgcggtcactaacgtcacgatacggcaacctgcggctccggct 391
            |  | ||||| |  |||  ||| ||||| || ||  | || || ||||| ||  |
Db    450 gtgctgagcggtagtggtttggttactaatgttactctgcgtcagcctgctgcatc---- 505

Qy    392 ggtggaggtgtgattaccctgcatggtcggtttgacattttgtctttgaccggtactgcg 451
            || || ||  |||   | | ||||| |||||| || |||||| ||   ||  ||
Db    506 --cggtggagttgttagtttacgtggtcagtttgagatcttgtctatgtgtggggctttt 563

Qy    452 cttccaccgcctg---caccaccgggagcaggaggtttgacggtgtatctagccggaggt 508
          |||||  || |||   | ||  | | ||| |  ||||| ||  | ||  |||| |||| |
Db    564 cttcctacgtctggctctcctgctgcagccgctggtttaaccatttacttagctggagct 623

Qy    509 caaggacaagttgtaggagggaatgtggctggttcgttaattgcttcgggaccggtagtg 568
          ||||| |||||||| |||||    || |||||  || | ||||| || ||||| ||  |
Db    624 caaggtcaagttgtgggaggtggagttgctggcccgcttattgcctctggacccgttatt 683

Qy    569 ttgatggctgcttcttttgcaaacgcagtttatgataggttaccgattgaagaggaagaa 628
            |||| ||||||  | |||   || ||   |||||| |||||||||||||| || ||| ||
Db    684 gtgatagctgctacgttttgcaatgccacttatgagaggttaccgattgaggaagaacaa 743

Qy    629 accccaccgccg 640
            | |  || |
Db    744 cagcaagagcag 755
```

RESULT 9
AAS96327
ID AAS96327 standard; cDNA; 1251 BP.
XX
AC AAS96327;
XX
DT 26-FEB-2002 (first entry)
XX
DE Arabidopsis cDNA encoding cell cycle protein CCP27.
XX
KW Cell cycle protein; CCP; ss; cell cycle regulation; herbicide;
KW plant growth regulator; plant development; abiotic stress; biotic stress;
KW nutrient deprivation; pathogen attack; crop yield.
XX
OS Arabidopsis thaliana.
XX
PN WO200185946-A2.
XX
PD 15-NOV-2001.
XX
PF 14-MAY-2001; 2001WO-IB01307.

```
XX
PR   12-MAY-2000; 2000US-204045P.
XX
PA   (CROP-) CROPDESIGN NV.
XX
PI   Inze D,  Boudolf V,  De Veylder L,  Acosta JAT,  Magyar Z;
XX
DR   WPI; 2002-062249/08.
DR   P-PSDB; AAU72537.
XX
PT   New cell cycle protein and nucleic acid molecule encoding it useful for
PT   regulating cell cycle progression in plants and for identifying
PT   modulators which are useful as herbicides or plant growth regulators -
XX
PS   Claim 38; Fig 31; 316pp; English.
XX
CC   The invention relates to a novel cell cycle protein (CCP) and the
CC   polynucleotides encoding them. CCP is useful for identifying a compound
CC   which modulates the activity of the polypeptide and which binds to the
CC   polypeptide and an anti-CCP antibody is useful for detecting the presence
CC   of CCP in a sample. A CCP modulator is useful for modulating the cell
CC   cycle or growth of a plant such as Arabidopsis thaliana, rice, wheat,
CC   maize, tomato, alfalfa, oilseed rape, soybean, sunflower and canola.
CC   CCP nucleic acid and polypeptide molecules are useful as modulating
CC   agents in regulating cell cycle progression in plants. CCP is useful to
CC   treat disorders characterised by insufficient or excessive production of
CC   CCP protein or production of CCP protein forms which have decreased or
CC   aberrant activity. Compounds that bind to or modulate the activity
CC   of CCP polypeptide are useful as herbicides or plant growth regulators.
CC   The polynucleotide is useful for modifying cell fate, plant development,
CC   plant morphology, biochemistry and/or physiology, the length of the G1,
CC   S, G2 and/or M phase of the cell cycle of a plant, initiation, promotion,
CC   stimulation or enhancement of cell division, DNA replication, seed set,
CC   seed size, seed development, tuber, fruit, leaf formation, shoot and root
CC   initiation and/or development, nodule function, dwarfism in plants,
CC   senescence, tolerance or resistance to stress. CCP, the polynucleotide
CC   and the anti-CCP antibody are useful in agriculture to modulate the
CC   protein levels or activity of a protein involved in the cell cycle due
CC   to environmental conditions, including abiotic stress such as
CC   cold, nutrient deprivation, heat, drought, salt stress, or biotic
CC   stress such as pathogen attack, to modulate e.g. enhance crop yields,
CC   and attenuate plant architecture, plant quality traits, plant
CC   reproduction and seed development, endoreduplication in storage cells,
CC   storage tissues and/or storage organs of plants or its parts. CCP is
CC   useful as an immunogen to generate antibodies. CCP protein is useful to
CC   screen for naturally occurring CCP substrates. The polynucleotide is
CC   useful for expressing CCP protein, to detect CCP mRNA, or a genetic
CC   lesion in a CCP gene and to modulate CCP activity. The present sequence
CC   encodes a CCP protein of the invention.
XX
SQ   Sequence 1251 BP; 301 A; 254 C; 335 G; 361 T; 0 other;

  Query Match             18.4%;  Score 178.8;  DB 24;  Length 1251;
  Best Local Similarity   63.2%;  Pred. No. 2.9e-46;
  Matches  311;  Conservative    0;  Mismatches  172;  Indels    9;  Gaps    2;

Qy    152 tcctccggacgtcgtccacgtggtcgtcctgcaggttccaaaaacaaacccaaacctccg 211
          ||  | || ||||||||  | ||| | ||| | |||||||| |||||||| ||   |||
Db    283 tctacgggtcgtcgtcctagaggtagacctcctggttccaagaacaaaccaaagagtcca 342

Qy    212 acgattataactagagatagtcctaacgtccttagatcacacgttcttgaagtcacctcc 271
            || | || | ||| || |||||   ||       || ||||||||  |  |  |
Db    343 gttgttgttaccaaagaaagccctaactctctccagagccatgttcttgagattgctacg 402

Qy    272 ggttcggacatatccgaggcagtctccacctacgccactcgtcgcggctgcggcgtttgc 331
          ||  | ||| |  | ||     |   | |||  ||   | | |||||| | |||||||
Db    403 ggagctgacgtggcggaaagcttaaacgcctttgctcgtagacgcggccggggcgtttcg 462

Qy    332 attataagcggcacgggtgcggtcactaacgtcacgatacggcaacctgcggctccggct 391
            |  | ||||| |  |||  ||| ||||| || ||  | || || ||||| ||  |
Db    463 gtgctgagcggtagtggtttggttactaatgttactctgcgtcagcctgctgcatc---- 518

Qy    392 ggtggaggtgtgattaccctgcatggtcggtttgacattttgtctttgaccggtactgcg 451
          || || ||  |||   | | ||||| |||||| || |||||| ||   ||  ||
Db    519 --cggtggagttgttagtttacgtggtcagtttgagatcttgtctatgtgtggggctttt 576

Qy    452 cttccaccgcctg---caccaccgggagcaggaggtttgacggtgtatctagccggaggt 508
          |||||  || |||   | ||  | | ||| |  ||||| ||  | ||  |||| |||| |
Db    577 cttcctacgtctggctctcctgctgcagccgctggtttaaccatttacttagctggagct 636

Qy    509 caaggacaagttgtaggagggaatgtggctggttcgttaattgcttcgggaccggtagtg 568
          ||||| |||||||| |||||    || |||||  || | ||||| || |||||  ||  |
Db    637 caaggtcaagttgtgggaggtggagttgctggcccgcttattgcctctggacccgttatt 696

Qy    569 ttgatggctgcttcttttgcaaacgcagtttatgataggttaccgattgaagaggaagaa 628
            |||| |||||| | |||   || ||   |||||| ||||||||||||||| || ||| ||
Db    697 gtgatagctgctacgttttgcaatgccacttatgagaggttaccgattgaggaagaacaa 756

Qy    629 accccaccgccg 640
            | |  || |
Db    757 cagcaagagcag 768


RESULT  10
AAC35116
ID   AAC35116 standard; DNA; 1303 BP.
XX
AC   AAC35116;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 9053.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
```

PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.

PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 18.2%; Score 177.2; DB 21; Length 1303;
Best Local Similarity 61.8%; Pred. No. 9.6e-46;
Matches 309; Conservative 0; Mismatches 173; Indels 18; Gaps 1;

```
Qy    153 cctccggacgtcgtccacgtggtcgtcctgcaggttccaaaaacaaacccaaacctccga 212
          || ||   || ||||||||||| |||||||| |||||||| || || ||||
Db    457 cccccacgcgccgtccacgtggacgtcctgctggttccaagaacaaaccaaagccaccga 516

Qy    213 cgattataactagagatagtcctaacgtccttagatcacacgttcttgaagtcacctccg 272
          |  | ||| | |||  ||| || |  || |     || ||  | ||  || |    |
Db    517 tcttcgtcactcgcgattctccaaatgctctcaagagccatgtcatggagatcgctagtg 576

Qy    273 gttcggacatatccgaggcagtctccacctacgccactcgtcgcggctgcggcgtttgca 332
          |  | ||| |   |||  |  |  | || |  || |  || ||     | ||| | ||||
Db    577 ggactgacgtcatcgaaaccctagctacttttgctaggcggcgtcaacgtggcatctgca 636

Qy    333 ttataagcggcacgggtgcggtcactaacgtcacgatacggcaac--------------- 377
          |  | |||||  |  ||  | ||  ||||||||||  | || ||||
Db    637 tcttgagcggaaatggcacagtggctaacgtcaccctccgtcaaccctcgaccgctgccg 696

Qy    378 ---ctgcggctccggctggtggaggtgtgattaccctgcatggtcggtttgacattttgt 434
             | |||||||| | |||||  | |||  |  |  | || ||  |||||||  ||| | |
Db    697 ttgcggcggctcctggtggtgcggctgttttggctttacaagggaggtttgagattcttt 756

Qy    435 ctttgaccggtactgcgcttccaccgcctgcaccaccgggagcaggaggtttgacggtgt 494
          ||||  ||||||  ||    | |||   || || |||||  ||   |    |||||  ||| | |
Db    757 ctttaaccggttctttcttgccaggaccggctccacctggttccaccggtttaacgattt 816

Qy    495 atctagccggaggtcaaggacaagttgtaggagggaatgtggctggttcgttaattgctt 554
          |  |||||||  |||||||| || ||||| |||||  |   ||||  || . | || || ||
Db    817 acttagccggtggtcaaggtcaggttgttggaggaagcgtggtgggcccattgatggcag 876

Qy    555 cgggaccggtagtgttgatggctgcttcttttgcaaacgcagtttatgataggttaccga 614
          | || |||||  || |||| || ||  |  ||  | |||||   ||| || || || ||
Db    877 caggtccggtgatgctgatcgccgccacgttctctaacgcgacttacgagagattgccat 936

Qy    615 ttgaagaggaagaaacccca 634
          | || ||||||||  |  ||
Db    937 tggaggaggaagaggcagca 956
```

RESULT 11
AAZ89379/c
ID AAZ89379 standard; DNA; 12319 BP.
XX
AC AAZ89379;
XX
DT 15-JUN-2000 (first entry)
XX
DE A. thaliana ACW1 genomic DNA.
XX
KW ACW1; plant cell wall; digestion efficiency; crop; agriculture; ss.
XX
OS Arabidopsis thaliana.
XX
PN JP2000041685-A.
XX
PD 15-FEB-2000.
XX
PF 23-MAR-1999; 99JP-0077502.
XX
PR 29-MAY-1998; 98JP-0166174.
XX
PA (OJIP) OJI PAPER CO.
PA (KAZU-) ZH KAZUSA DNA KENKYUSHO.
XX
DR WPI; 2000-274045/24.
DR P-PSDB; AAY51874.
XX
PT Modification of the cell wall components of a plant - useful for
PT improving digestion efficiency of crops and feed crops
XX
PS Disclosure; Page 14-20; 24pp; Japanese.
XX
CC This invention describes a novel nucleic acid used for modifying the
CC cell wall components of a plant. The modification is used for the
CC improvement in digestion efficiency of useful crops and feed crops.
CC The DNA is useful in agriculture, industry and gardening. This sequence
CC encodes the Arabidopsis thaliana ACW1 protein which is described in the
CC method of the invention.
XX
SQ Sequence 12319 BP; 3722 A; 2213 C; 2413 G; 3971 T; 0 other;

Query Match 13.1%; Score 127.6; DB 21; Length 12319;
Best Local Similarity 54.3%; Pred. No. 2.3e-29;
Matches 290; Conservative 0; Mismatches 229; Indels 15; Gaps 1;

```
Qy    161 cgtcgtccacgtggtcgtcctgcaggttccaaaaacaaacccaaacctccgacgattata 220
          || |||||||||||  | ||  |||||||||||||||||||| ||||| ||      |  |
Db  10851 CGACGTCCACGTGGCAGACCACCAGGTTCCAAAAACAAACCTAAACCACCCGTCTTCGTC 10792

Qy    221 actagagatagtcctaacgtccttagatcacacgttcttgaagtcacctccggttcggac 280
          ||  | || |             | ||  |  || | || |||||  | || ||    |||
Db  10791 ACACGTGACACCGACCCTCCTATGAGTCCTTACATCCTCGAAGTTCCTTCAGGAAACGAC 10732

Qy    281 atatccgaggcagtctccacctacgccactcgtcgcggctgcggcgtttgcattataagc 340
            |   ||| ||  ||  |   | |  |   |||     |  ||| || ||| |  | ||
Db  10731 GTCGTCGAAGCCATCAACCGTTTCTGCCGCCGTAAATCCATCGGAGTCTGCGTCCTTAGT 10672

Qy    341 ggcacgggtgcggtcactaacgtcacgatacggcaacctgcggctccggctggtggaggt 400
          ||| | ||  | ||  ||||||||||  |||| || ||  |  |  | |||  |||   |
Db  10671 GGCTCTGGCTCTGTAGCTAACGTCACTTTACGTCAGCCATCACCGGCAGCTCTTGGCTCT 10612

Qy    401 gtgattaccctgcatggtcggtttgacattttgtctttgaccggtactgcgcttccaccg 460
            || ||  | |||||   ||||||  |  | ||  |  |||  ||    || || ||
Db  10611 ACCATAACTTTCCATGGAAAGTTTGATCTCCTCTCCGTCTCCGCAACGTTTCTCCCTCCT 10552

Qy    461 cc---------------tgcaccaccgggagcaggaggtttgacggtgtatctagccgga 505
          ||                | || ||  |   |       || || || | ||| || |||
Db  10551 CCGCCTCGTACTTCCTTGTCTCCTCCCGTTTCTAACTTCTTCACCGTCTCTCTCGCTGGA 10492

Qy    506 ggtcaaggacaagttgtaggagggaatgtggctggttcgttaattgcttcgggaccggta 565
            |||||||||  |  | ||||||   || ||||||  |  | |||  |  | ||| | ||
Db  10491 CCTCAAGGACAAATCATCGGAGGGTTCGTCGCTGGTCCACTTATTTCGGCAGGAACAGTT 10432

Qy    566 gtgttgatggctgcttcttttgcaaacgcagtttatgataggttaccgattgaagaggaa 625
            | || || ||   |||   ||| |   |||| |  ||||||||   ||||| ||
Db  10431 TACGTCATCGCCGCAAGTTTCAACAACCCTTCTTATCACCGGTTACCGGCGGAAGAAGAG 10372

Qy    626 gaaaccccaccgccgagaaccaccggggtgcagcagcagcagccggaggcgtct 679
            |||  |   || || | ||    |       ||   ||    |||  ||  |||
Db  10371 CAAAAACACTCGGCGGGGACAGGGGAAAGAGAGGGACAATCTCCGCCGGTCTCT 10318
```

RESULT 12
AAC44421
ID AAC44421 standard; DNA; 1219 BP.
XX
AC AAC44421;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 42770.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX

PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.

```
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             4.3%;  Score 42;  DB 21;  Length 1219;
  Best Local Similarity   53.4%;  Pred. No. 0.013;
  Matches  143;  Conservative    0;  Mismatches  110;  Indels   15;  Gaps    2;

Qy     327 tttgcattataagcggcacgggtgcggtcactaacgtcacgatacggcaacctgcggctc 386
           ||||| || || |||  |  ||||  ||  | | ||| ||  | || || ||||   |
Db     747 tttgcgttttatgcgcaaacggtgtcgtttcgagcgttacacttcgtcagcctgattcat 806

Qy     387 cggctggtggaggtgtgattaccctgcatggtcggtttgacattttgtctttgaccggta 446
           |    ||| |||    | |||    | ||||||||||| || |||||  |  | || |
Db     807 c------tggtggtacattgacctatgagggtcggtttgagatattgtcactatctggaa 860

Qy     447 ctgcgcttc---------caccgcctgcaccaccgggagcaggaggtttgacggtgtatc 497
           |     | |         ||       ||| |    || |||| ||  |||  ||||  |
Db     861 cattcatgcctagtgactcagacgggacacgaagcagaacaggcgggatgagcgtgtcgc 920

Qy     498 tagccggaggtcaaggacaagttgtaggagggaatgtggctggttcgttaattgcttcgg 557
           | ||  |   | | ||||  || ||||| ||   ||| ||||| | | |  ||||  |
Db     921 ttgctagccctgatggacgtgtagtaggtggtggtgttgctggcttgctggttgcagcca 980

Qy     558 gaccggtagtgttgatggctgcttctttt 585
             ||  |    || | |  |   ||||
Db     981 ctcctattcaagtggttgtaggaacttt 1008


RESULT  13
AAD05779
ID   AAD05779 standard; cDNA; 1263 BP.
XX
AC   AAD05779;
XX
DT   31-JUL-2001  (first entry)
XX
DE   Arabidopsis thaliana transcription factor, G605 cDNA.
XX
KW   Transcription factor; biochemical characteristic; controlling element;
KW   structural characteristic; developmental characteristic; gene therapy;
KW   agricultural biotechnology; plant trait modification; ss.
XX
OS   Arabidopsis thaliana.
XX
FH   Key             Location/Qualifiers
FT   CDS             72..1076
FT                   /*tag= a
FT                   /product= "Transcription factor, G605"
XX
PN   WO200136597-A1.
XX
PD   25-MAY-2001.
XX
PF   14-NOV-2000; 2000WO-US31344.
XX
PR   17-NOV-1999;   99US-0166228.
PR   17-APR-2000; 2000US-0197899.
PR   22-AUG-2000; 2000US-0227439.
XX
PA   (MEND-) MENDEL BIOTECHNOLOGY INC.
PA   (CREE/) CREELMAN R.
PA   (YUGG/) YU G.
PA   (ADAM/) ADAM L.
PA   (RIEC/) RIECHMANN J L.
PA   (HEAR/) HEARD J.
PA   (SAMA/) SAMAHA R.
PA   (PILG/) PILGRIM M.
PA   (PINE/) PINEDA O.
PA   (JIAN/) JIANG C.
XX
PI   Creelman R,  Yu G,  Adam L,  Riechmann JL,  Heard J,  Samaha R;
PI   Pilgrim M,  Pineda O,  Jiang C;
XX
DR   WPI; 2001-335999/35.
DR   P-PSDB; AAE01895.
XX
PT   Nucleic acids encoding plant transcription factor polypeptides, useful
PT   for altering the biochemical characteristics of plants e.g. corn,
PT   potato and cotton plants -
XX
PS   Claim 4; Page 75-76; 127pp; English.
XX
CC   The present sequence is Arabidopsis thaliana transcription factor,
CC   G605 cDNA. The transcription factor is used for altering a plant's
CC   biochemical characteristics. The transcription factor may be used to
CC   alter the structure and developmental characteristics of plants such as
CC   soybean, wheat, corn, potato, cotton, rice, oilseed rape, sunflower,
CC   alfalfa, sugar cane, turf, banana, blackberry, blueberry, strawberry,
CC   raspberry, cantaloupe, carrot, cauliflower, coffee, cucumber, eggplant,
CC   grapes, honey dew, lettuce, mango, melon, onion, papaya, peas, peppers,
CC   pineapple, spinach, squash, sweet corn, tobacco, tomato, watermelon,
CC   rosaceous fruits and/or vegetable brassicas. Transcription factors are
CC   key controlling elements of biological pathways and altering expression
CC   levels of 1 or more transcription factors can change entire biological
CC   pathways in an organism. Therefore manipulating transcription factor
CC   levels in plants offers great potential in agricultural biotechnology
CC   for modifying a plant's traits. Transcription factor cDNA is useful in
CC   gene therapy.
XX
SQ   Sequence 1263 BP; 335 A; 314 C; 274 G; 340 T; 0 other;


  Query Match             4.3%;  Score 42;  DB 22;  Length 1263;
  Best Local Similarity   53.4%;  Pred. No. 0.013;
  Matches  143;  Conservative    0;  Mismatches  110;  Indels   15;  Gaps    2;

Qy     327 tttgcattataagcggcacgggtgcggtcactaacgtcacgatacggcaacctgcggctc 386
           ||||| || || |||  |  ||||  ||  | | ||| ||  | || || ||||   |
```

```
Db    604 tttgcgttttatgcgcaaacggtgtcgtttcgagcgttacacttcgtcagcctgattcat 663

Qy    387 cggctggtggaggtgtgattaccctgcatggtcggtttgacattttgtctttgaccggta 446

Db    664 c------tggtggtacattgacctatgagggtcggtttgagatattgtcactatctggaa 717

Qy    447 ctgcgcttc---------caccgcctgcaccaccgggagcaggaggtttgacggtgtatc 497

Db    718 cattcatgcctagtgactcagacgggacacgaagcagaacaggcgggatgagcgtgtcgc 777

Qy    498 tagccggaggtcaaggacaagttgtaggagggaatgtggctggttcgttaattgcttcgg 557

Db    778 ttgctagccctgatggacgtgtagtaggtggtggtgttgctggcttgctggttgcagcca 837

Qy    558 gaccggtagtgttgatggctgcttcttt 585

Db    838 ctcctattcaagtggttgtaggaacttt 865
```

RESULT 14
AAC38928
ID AAC38928 standard; DNA; 1387 BP.
XX
AC AAC38928;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 22725.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.

PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 4.3%; Score 42; DB 21; Length 1387;
Best Local Similarity 53.4%; Pred. No. 0.014;
Matches 143; Conservative 0; Mismatches 110; Indels 15; Gaps 2;

```
Qy    327 tttgcattataagcggcacgggtgcggtcactaacgtcacgatacggcaacctgcggctc 386
          ||||| || || |||  |  ||||  ||  | | ||| ||  | || || ||||   |
Db    747 tttgcgttttatgcgcaaacggtgtcgtttcgagcgttacacttcgtcagcctgattcat 806

Qy    387 cggctggtggaggtgtgattaccctgcatggtcggtttgacattttgtctttgaccggta 446
          |      ||| |||    | |||    | ||||||||||| || |||||  |  | || |
Db    807 c------tggtggtacattgacctatgagggtcggtttgagatattgtcactatctggaa 860

Qy    447 ctgcgcttc---------caccgcctgcaccaccgggagcaggaggtttgacggtgtatc 497
          |    | |         ||       ||| |    || |||| ||  |||  ||||  |
Db    861 cattcatgcctagtgactcagacgggacacgaagcagaacaggcgggatgagcgtgtcgc 920

Qy    498 tagccggaggtcaaggacaagttgtaggagggaatgtggctggttcgttaattgcttcgg 557
          | ||  |   | | ||||  || ||||| ||   ||| ||||| | | |  |||| |
Db    921 ttgctagccctgatggacgtgtagtaggtggtggtgttgctggcttgctggttgcagcca 980

Qy    558 gaccggtagtgttgatggctgcttcttt 585
            ||  |     || | |  |   ||||
Db    981 ctcctattcaagtggttgtaggaacttt 1008
```

RESULT 15
AAD06652
ID AAD06652 standard; cDNA; 1606 BP.
XX
AC AAD06652;
XX
DT 10-AUG-2001 (first entry)
XX
DE A. thaliana transcription factor G1068 cDNA.
XX
KW Plant transcription factor; phenotype; sugar sensing characteristic;
KW transgenic plant; plant yield; growth; germination; photosynthesis;
KW glyoxylate metabolism; respiration; pathogen response; wounding response;
KW cell cycle regulation; pigmentation; flowering; senescence; physiology;
KW storage organ; metabolism; ss.
XX
OS Arabidopsis thaliana.
XX
FH Key Location/Qualifiers
FT CDS 150..1310
FT /*tag= a
FT /product= "Transcription factor"
XX
PN WO200135725-A1.
XX
PD 25-MAY-2001.
XX
PF 14-NOV-2000; 2000WO-US31414.
XX
PR 17-NOV-1999; 99US-0166228.
PR 17-APR-2000; 2000US-0197899.
PR 22-AUG-2000; 2000US-0227439.
XX
PA (MEND-) MENDEL BIOTECHNOLOGY INC.
PA (JIAN/) JIANG C.
PA (HEAR/) HEARD J.
PA (PINE/) PINEDA O.
PA (PILG/) PILGRIM M.
PA (ADAM/) ADAM L.
PA (RIEC/) RIECHMANN J L.
PA (YUGG/) YU G.
PA (SAMA/) SAMAHA R.
XX
PI Jiang C, Heard J, Pineda O, Pilgrim M, Adam L, Riechmann JL;
PI Yu G, Samaha R;

XX
DR WPI; 2001-335977/35.
DR P-PSDB; AAE02551.
XX
PT Nucleic acids encoding plant transcription factor polypeptides, useful
PT for altering the sugar sensing characteristics of plants and increasing
PT yield, e.g. corn, potato and cotton plants -
XX
PS Claim 4; Page 90-92; 151pp; English.
XX
CC The patent relates to polynucleotides encoding 35 plant transcription
CC factors which may be used to modify phenotype associated with a plant's
CC sugar sensing characteristics and increasing yield when their expression
CC level is altered. Sugars are central regulatory molecules that control
CC aspects of physiology, metabolism and development. Therefore the cDNAs
CC and proteins of the invention are useful for modifying the growth and
CC germination rates of plants, photosynthesis, glyoxylate metabolism,
CC respiration, starch and sucrose synthesis and degradation, pathogen
CC response, wounding response, cell cycle regulation, pigmentation,
CC flowering and senescence of plants and for modifying sink-source
CC relationships in seeds, tubers, roots, and other storage organs leading
CC to an increase in yield. The transcription factor polynucleotides and
CC polypeptides may be used to alter the structure and developmental
CC characteristics of plants such as soybean, wheat, corn, potato, cotton,
CC rice, oilseed, sunflower, alfalfa, sugarcane, turf, banana, blackberry,
CC blueberry, strawberry, raspberry, cantaloupe, carrot, cauliflower,
CC coffee, cucumber, eggplant, grapes, honey dew, lettuce, mango, melon,
CC onion, papaya, peas, peppers, pineapple, spinach, squash, sweet corn,
CC tobacco, tomato, watermelon, rosaceous fruits and/or vegetable brassicas.
CC The present sequence is an Arabidopsis thaliana transcription factor
CC cDNA.
XX
SQ Sequence 1606 BP; 459 A; 376 C; 335 G; 436 T; 0 other;

Query Match 4.3%; Score 42; DB 22; Length 1606;
Best Local Similarity 47.4%; Pred. No. 0.015;
Matches 166; Conservative 0; Mismatches 175; Indels 9; Gaps 1;

```
Qy    249 cacacgttcttgaagtcacctccggttcggacatatccgaggcagtctccacctacgcca 308
          | || ||  |||| || |   | ||   |||||||  |   |    | |   | |   | |
Db    658 ctcatgtcattgaggttaaaacaggagaggacatagctactgcgaagatattggcgtttacga 717

Qy    309 ctcgtcgcggctgcggcgtttgcattataagcggcacgggtgcggtcactaacgtcacga 368
            |   |     |||   | || ||| |    |  || || || || ||||| || | |
Db    718 accaagggccacgcgcaatctgtattctctcagctacaggagctgtaactaatgtgatgc 777

Qy    369 tacggcaacctgcggctccggctggtggaggtgtgattaccctgcatggtcggtttgaca 428
          | || ||| ||     |    |  |   ||     |||      | || || ||||| |
Db    778 ttcgtcaagctaacaatagcaatcctactggaactgttaagtatgagggccgatttgaaa 837

Qy    429 ttttgtctttgaccggtactgcgcttccaccgcc---------tgcaccaccgggagcag 479
          |  | ||| || | ||| ||    |     |              | |     |   | | |
Db    838 tcatttctctgtcaggttctttcttgaattctgagagtaatggtactgtgaccaaaactg 897

Qy    480 gaggtttgacggtgtatctagccggaggtcaaggacaagttgtaggagggaatgtggctg 539
          |    ||||  ||||  || || |||    |||| |   |||| || ||   ||| | ||
Db    898 gtaacttgagtgtgtcgctggctggacacgaaggccggattgtgggtggatgtgttgatg 957

Qy    540 gttcgttaattgcttcgggaccggtagtgttgatggctgcttcttttgca 589
          |   | || | |||    ||  ||   | | || |  |     |||| |
Db    958 gaatgctagtagctggatcacaagtccaggtcattgtgggaagctttgta 1007
```

Search completed: July 10, 2002, 22:14:29
Job time: 7602 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2002, 19:59:17 ; Search time 63.21 Seconds
(without alignments)
3784.955 Million cell updates/sec

Title: US-09-823-676-1
Perfect score: 974
Sequence: 1 ccccccgacctgcctctaca.....cgaaatgcgaatattaggtt 974

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	44.6	4.6	7218	1	US-08-232-463-14	Sequence 14, Appl
	2	33.8	3.5	8868	3	US-08-815-809-7	Sequence 7, Appli
	3	33.8	3.5	19877	2	US-08-816-155B-8	Sequence 8, Appli
	4	33.8	3.5	19877	3	US-09-079-587-8	Sequence 8, Appli
c	5	33.6	3.4	2793	1	US-08-209-747-1	Sequence 1, Appli
c	6	33.6	3.4	2793	1	US-08-458-298-1	Sequence 1, Appli
c	7	33.6	3.4	9468	1	US-08-325-547-10	Sequence 10, Appl
c	8	33.4	3.4	1100	3	US-08-642-807A-32	Sequence 32, Appl
c	9	31.8	3.3	4403765	4	US-09-103-840A-2	Sequence 2, Appli
	10	31.6	3.2	5433	3	US-08-929-329-1	Sequence 1, Appli
c	11	31.6	3.2	22671	4	US-08-976-259-14	Sequence 14, Appl
c	12	31.4	3.2	765	4	US-09-199-637A-324	Sequence 324, App
c	13	31.4	3.2	1071	4	US-09-199-637A-322	Sequence 322, App
	14	31.4	3.2	1590	4	US-09-199-637A-320	Sequence 320, App
c	15	31.4	3.2	42235	4	US-09-199-637A-1	Sequence 1, Appli
	16	31.2	3.2	720	4	US-08-998-416-859	Sequence 859, App
c	17	31.2	3.2	1028	4	US-08-118-200-1	Sequence 1, Appli
c	18	31.2	3.2	1028	4	US-08-458-745-1	Sequence 1, Appli
	19	31.2	3.2	6085	4	US-09-029-603-4	Sequence 4, Appli
	20	31.2	3.2	80246	4	US-09-078-294-4	Sequence 4, Appli
	21	31	3.2	2790	2	US-08-718-661-1	Sequence 1, Appli
c	22	31	3.2	6306	1	US-08-466-390-3	Sequence 3, Appli
c	23	31	3.2	6306	1	US-08-470-950-3	Sequence 3, Appli
c	24	31	3.2	6306	1	US-08-467-781-3	Sequence 3, Appli
c	25	31	3.2	6306	1	US-08-195-487-3	Sequence 3, Appli
c	26	31	3.2	6306	2	US-08-483-924-3	Sequence 3, Appli
c	27	31	3.2	6306	5	PCT-US93-06160-3	Sequence 3, Appli
	28	31	3.2	7218	1	US-08-232-463-14	Sequence 14, Appl
	29	30.8	3.2	407	4	US-09-056-556-173	Sequence 173, App
	30	30.8	3.2	1683	1	US-07-945-283-3	Sequence 3, Appli
c	31	30.8	3.2	1831	6	5215881-1	Patent No. 5215881
c	32	30.8	3.2	1831	6	5215881-3	Patent No. 5215881
c	33	30.8	3.2	8438	1	US-07-945-283-1	Sequence 1, Appli
c	34	30.6	3.1	3915	2	US-08-485-139-6	Sequence 6, Appli
c	35	30.6	3.1	3915	3	US-08-750-357-6	Sequence 6, Appli
c	36	30.6	3.1	4137	3	US-08-750-357-7	Sequence 7, Appli
c	37	30.6	3.1	31571	1	US-08-323-443B-1	Sequence 1, Appli
c	38	30.6	3.1	53526	3	US-08-658-136-2	Sequence 2, Appli
c	39	30.6	3.1	53577	3	US-08-658-136-1	Sequence 1, Appli
	40	30.4	3.1	396	2	US-08-244-205-14	Sequence 14, Appl
	41	30.4	3.1	396	5	PCT-US92-10284-14	Sequence 14, Appl
c	42	30.4	3.1	3541	4	US-08-976-259-80	Sequence 80, Appl
c	43	30.4	3.1	6530	2	US-08-146-930-1	Sequence 1, Appli
c	44	30.4	3.1	6530	3	US-08-458-240-1	Sequence 1, Appli
c	45	30.4	3.1	6530	5	PCT-US93-03993-1	Sequence 1, Appli

ALIGNMENTS

```
RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-Fls
US-08-232-463-14
```

```
  Query Match             4.6%;  Score 44.6;  DB 1;  Length 7218;
  Best Local Similarity   3.5%;  Pred. No. 0.0045;
  Matches   14;  Conservative  216;  Mismatches  165;  Indels    0;  Gaps    0;

Qy     473 ggagcaggaggtttgacggtgtatctagccggaggtcaaggacaagttgtaggagggaat 532
           ||  |:::::   :: :: : :   ::  :::::  ::::: :::  : :::::::::
Db    1439 GGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1380

Qy     533 gtggctggttcgttaattgcttcgggaccggtagtgttgatggctgcttcttttgcaaac 592
           : ::  ::   :  ::  :    ::::  :: :: :  :: ::  :         : :::
Db    1379 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1320

Qy     593 gcagtttatgataggttaccgattgaagaggaagaaaccccaccgccgagaaccaccggg 652
           : ::   : :: :::  :  ::  :::::::::::::    :  :  :::::  :  :::
Db    1319 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1260

Qy     653 gtgcagcagcagcagccggaggcgtctcagtcgtcggaggttacggggagtggggcccag 712
           : : :: :: :: ::  ::::: :    ::  :  :::::  : :::::: ::::   ::
Db    1259 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1200

Qy     713 gcgtgtgagtcaaacctccaaggtggaaatggtggaggaggtgttgctttctacaatctt 772
           : : : :::  :::     :::: ::::: :: :::::::: :  :       : ::
Db    1199 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1140

Qy     773 ggaatgaatatgaacaattttcaattctccgggggagatatttacggtatgagcggcggt 832
           :::: ::: : ::: ::     ::      :::::::: :   : :: : ::: :: ::
Db    1139 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1080

Qy     833 agcggaggaggtggtggcggtgcgactagacccgc 867
           :: :::::::: : | ||       || |||| ||
Db    1079 RRRRRRRRRRRRRATCGCAAGCTCCCTCGACCTGC 1045


RESULT   2
US-08-815-809-7
; Sequence 7, Application US/08815809
; Patent No. 6004777
; GENERAL INFORMATION:
;  APPLICANT: TARTAGLIA, James
;  APPLICANT: GOEBEL, Scott J.
;  APPLICANT: COX, William I.
;  APPLICANT: GETTIG, Russell R.
;  APPLICANT: PINCUS, Steven E.
;  APPLICANT: PAOLETTI, Enzo
;  APPLICANT: JACOBS, Bertram L.
;  TITLE OF INVENTION: VECTORS HAVING ENHANCED EXPRESSION, AND METHODS OF
;  TITLE OF INVENTION: MAKING AND USES THEREOF
;  FILE REFERENCE: 454310-3010
;  CURRENT APPLICATION NUMBER: US/08/815,809
;  CURRENT FILING DATE: 1997-03-12
;  NUMBER OF SEQ ID NOS: 23
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
;   LENGTH: 8868
;   TYPE: DNA
;   ORGANISM: Vaccinia virus
US-08-815-809-7


  Query Match             3.5%;  Score 33.8;  DB 3;  Length 8868;
  Best Local Similarity   51.7%;  Pred. No. 6.8;
  Matches   77;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy     483 gtttgacggtgtatctagccggaggtcaaggacaagttgtaggagggaatgtggctggtt 542
           ||   | |||||  | |   |  |||  |   | || || ||  |  || ||| | || ||
Db    7859 gtaaaatggtgtttattatcctaggagatctataaattataatatagattgttgttgttt 7918

Qy     543 cgttaattgcttcgggaccggtagtgttgatggctgcttcttttgcaaacgcagtttatg 602
             |||| |   |   ||||||  |  | |  | || || ||    |||   ||| |||
Db    7919 ttttaaattgctacataccggtcttcataaaagatgattattagataaagatagtctatc 7978

Qy     603 ataggttaccgattgaagaggaagaaacc 631
           |  |   |||   | || |   || |
Db    7979 ttatctacaagatatgaaagaagataatc 8007


RESULT   3
US-08-816-155B-8
; Sequence 8, Application US/08816155B
; Patent No. 5990091
;  GENERAL INFORMATION:
;    APPLICANT:  TARTAGLIA, JAMES
;    APPLICANT:  COX, WILLIAM I.
;    APPLICANT:  GETTIG, RUSSELL R.
;    APPLICANT:  MARTINEZ, HECTOR
;    APPLICANT:  PAOLETTI, ENZO
;    APPLICANT:  PINCUS, STEVEN E.
;    TITLE OF INVENTION:  VECTORS HAVING ENHANCED EXPRESSION, AND
;    TITLE OF INVENTION:  METHODS OF MAKING AND USES THEREOF
;    NUMBER OF SEQUENCES:  48
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FROMMER LAWRENCE & HAUG LLP
;      STREET:  745 FIFTH AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10151
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/816,155B
;      FILING DATE:  12-MAR-1997
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KOWALSKI, THOMAS J.
;      REGISTRATION NUMBER:  32,147
;      REFERENCE/DOCKET NUMBER:  454310-2990
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-588-0800
;      TELEFAX:  212-588-0500
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  19877 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-816-155B-8


  Query Match             3.5%;  Score 33.8;  DB 2;  Length 19877;
  Best Local Similarity   51.7%;  Pred. No. 10;
  Matches   77;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy     483 gtttgacggtgtatctagccggaggtcaaggacaagttgtaggagggaatgtggctggtt 542
           ||   | |||||  | |   |  |||  |   | || || ||  |  || ||| | || ||
Db    7909 GTAAAATGGTGTTTATTATCCTAGGAGATCTATAAATTATAATATAGATTGTTGTTGTTT 7968

Qy     543 cgttaattgcttcgggaccggtagtgttgatggctgcttcttttgcaaacgcagtttatg 602
             |||| |   |   ||||||  |  | |  | || || ||    |||   ||| |||
Db    7969 TTTTAAATTGCTACATACCGGTCTTCATAAAAGATGATTATTAGATAAAGATAGTCTATC 8028

Qy     603 ataggttaccgattgaagaggaagaaacc 631
           |  |   |||   | || |   || |
Db    8029 TTATCTACAAGATATGAAAGAAGATAATC 8057


RESULT   4
US-09-079-587-8
; Sequence 8, Application US/09079587
; Patent No. 6130066
;  GENERAL INFORMATION:
;    APPLICANT:  TARTAGLIA, JAMES
```

```
;    APPLICANT:  COX, WILLIAM I.
;    APPLICANT:  GETTIG, RUSSELL R.
;    APPLICANT:  MARTINEZ, HECTOR
;    APPLICANT:  PAOLETTI, ENZO
;    APPLICANT:  PINCUS, STEVEN E.
;    TITLE OF INVENTION:  VECTORS HAVING ENHANCED EXPRESSION, AND
;    TITLE OF INVENTION:  METHODS OF MAKING AND USES THEREOF
;    NUMBER OF SEQUENCES:  48
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FROMMER LAWRENCE & HAUG LLP
;      STREET:  745 FIFTH AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10151
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/079,587
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/816,155
;      FILING DATE:  12-MAR-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KOWALSKI, THOMAS J.
;      REGISTRATION NUMBER:  32,147
;      REFERENCE/DOCKET NUMBER:  454310-2990
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-588-0800
;      TELEFAX:  212-588-0500
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  19877 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-079-587-8


  Query Match             3.5%;  Score 33.8;  DB 3;  Length 19877;
  Best Local Similarity   51.7%;  Pred. No. 10;
  Matches   77;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy     483 gtttgacggtgtatctagccggaggtcaaggacaagttgtaggagggaatgtggctggtt 542
           ||   | |||||  | |    |  |||  |   | || || ||  |  || ||| | || ||
Db    7909 GTAAAATGGTGTTTATTATCCTAGGAGATCTATAAATTATAATATAGATTGTTGTTGTTT 7968

Qy     543 cgttaattgcttcgggaccggtagtgttgatggctgcttcttttgcaaacgcagtttatg 602
             |||| |   |    ||||||  |  | |  | || || ||    |||   ||| |||
Db    7969 TTTTAAATTGCTACATACCGGTCTTCATAAAAGATGATTATTAGATAAAGATAGTCTATC 8028

Qy     603 ataggttaccgattgaagaggaagaaacc 631
           ||  |    |||   | || |   || |
Db    8029 TTATCTACAAGATATGAAAGAAGATAATC 8057


RESULT   5
US-08-209-747-1/c
; Sequence 1, Application US/08209747
; Patent No. 5733771
;  GENERAL INFORMATION:
;    APPLICANT:  Lewis, Randolph V.
;    APPLICANT:  Colgin, Mark
;    TITLE OF INVENTION:  cDNAs Encoding Minor Ampullate Spider
;    TITLE OF INVENTION:  Silk Proteins
;    NUMBER OF SEQUENCES:  56
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Birch, Stewart, Kolasch & Birch
;      STREET:  P.O. Box 747
;      CITY:  Falls Church
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22040-3487
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/209,747
;      FILING DATE:  14-MAR-1994
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murphy Jr., Gerald M.
;      REGISTRATION NUMBER:  28,977
;      REFERENCE/DOCKET NUMBER:  1447-104P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-205-8000
;      TELEFAX:  703-205-8050
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2793 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Nephila clavipes
;      TISSUE TYPE:  minor ampullate gland
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  183..2675
;      OTHER INFORMATION:  /product= "N. clavipes minor
;      OTHER INFORMATION:  ampullate silk protein"
US-08-209-747-1


  Query Match             3.4%;  Score 33.6;  DB 1;  Length 2793;
  Best Local Similarity   53.9%;  Pred. No. 4.4;
  Matches   69;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy     268 ctccggttcggacatatccgaggcagtctccacctacgccactcgtcgcggctgcggcgt 327
           | |||   |   | |  ||   | || ||||| || |  ||| ||  |||||| | ||
Db    1388 CACCGTACCCACCTTGACCACCGTAGCCTCCAGCTCCTGCACCCGCAGCGGCTCCTGCTC 1329

Qy     328 ttgcattataagcggcacgggtgcggtcactaacgtcacgatacggcaacctgcggctcc 387
             |||     |||||| |  |  |   ||| | |  | |       | ||| ||| ||||
Db    1328 CAGCACCCGCAGCGGCTCCTGCTCCAGCACCAGCACCTCACCGTAACCACCAGCGCCTCC 1269

Qy     388 ggctggtg 395
             |||  ||
Db    1268 AGCTCCTG 1261


RESULT   6
US-08-458-298-1/c
; Sequence 1, Application US/08458298
; Patent No. 5756677
;  GENERAL INFORMATION:
;    APPLICANT:  Lewis, Randolph V.
;    APPLICANT:  Colgin, Mark
;    TITLE OF INVENTION:  cDNAs Encoding Minor Ampullate Spider
;    TITLE OF INVENTION:  Silk Proteins
;    NUMBER OF SEQUENCES:  56
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Birch, Stewart, Kolasch & Birch
;      STREET:  P.O. Box 747
;      CITY:  Falls Church
;      STATE:  Virginia
;      COUNTRY:  USA
```

```
;      ZIP:  22040-3487
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,298
;      FILING DATE:  02-JUN-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/209,747
;      FILING DATE:  14-MAR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murphy Jr., Gerald M.
;      REGISTRATION NUMBER:  28,977
;      REFERENCE/DOCKET NUMBER:  1447-104P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-205-8000
;      TELEFAX:  703-205-8050
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2793 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Nephila clavipes
;      TISSUE TYPE:  minor ampullate gland
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  183..2675
;      OTHER INFORMATION:  /product= "N. clavipes minor
;      OTHER INFORMATION:  ampullate silk protein"
US-08-458-298-1


  Query Match             3.4%;  Score 33.6;  DB 1;  Length 2793;
  Best Local Similarity   53.9%;  Pred. No. 4.4;
  Matches   69;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy     268 ctccggttcggacatatccgaggcagtctccacctacgccactcgtcgcggctgcggcgt 327
           | |||   |   | |  ||   | || ||||| || |  ||| ||  |||||| | ||
Db    1388 CACCGTACCCACCTTGACCACCGTAGCCTCCAGCTCCTGCACCCGCAGCGGCTCCTGCTC 1329

Qy     328 ttgcattataagcggcacgggtgcggtcactaacgtcacgatacggcaacctgcggctcc 387
             |||     |||||| |  |  |   ||| | |  | |       | ||| ||| ||||
Db    1328 CAGCACCCGCAGCGGCTCCTGCTCCAGCACCAGCACCTCACCGTAACCACCAGCGCCTCC 1269

Qy     388 ggctggtg 395
           |||  ||
Db    1268 AGCTCCTG 1261


RESULT   7
US-08-325-547-10/c
; Sequence 10, Application US/08325547
; Patent No. 5736378
;  GENERAL INFORMATION:
;    APPLICANT:  Elder, John R.
;    APPLICANT:  Talbott, Randy L.
;    TITLE OF INVENTION:  NUCLEOTIDE SEQUENCE AND OPEN READING
;    TITLE OF INVENTION:  FRAMES OF FELINE IMMUNODEFICIENCY VIRUS (FIV)
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Spensley Horn Jubas & Lubitz
;      STREET:  1880 Century Park East, Suite 500
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  90067
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/325,547
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/759,570
;      FILING DATE:  12-SEP-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Wetherell, Jr., Ph.D., John R.
;      REGISTRATION NUMBER:  31,678
;      REFERENCE/DOCKET NUMBER:  PD-2000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 455-5100
;      TELEFAX:  (619) 455-5110
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9468 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    IMMEDIATE SOURCE:
;      CLONE:  P
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(628..1977, 1869..5240, 5236..5988, 5992
;      LOCATION:  ..6222, 6262..8824, 6709..6912)
US-08-325-547-10


  Query Match             3.4%;  Score 33.6;  DB 1;  Length 9468;
  Best Local Similarity   51.3%;  Pred. No. 8;
  Matches   78;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy     748 aggaggtgttgctttctacaatcttggaatgaatatgaacaattttcaattctccgggggg 807
           | || ||  |  |  ||||  | ||| |||  |||    ||||  |   | ||||   ||
Db    2010 ATGATGTAGTAGTACCTACTTTATTGTAATTTATAAATCCAATAATTTTTCCTCCATTGG 1951

Qy     808 agatatttacggtatgagcggcggtagcggaggaggtggtggcggtgcgactagacccgc 867
           || |    | || || |  |   |  ||    ||     ||| | |||  || | |||||
Db    1950 AGGTGCAGATGGCATTACTGTTTGCTGCACTTGATTCACTGGGGCTGCAGCTCGCCCCGC 1891

Qy     868 gttttagagttttagcgttttggtgacaccttt 899
           ||  ||  |      || |  |  ||||||
Db    1890 CTTCCAGTTTCCCGAATTTCTTTTACCACCTT 1859


RESULT   8
US-08-642-807A-32/c
; Sequence 32, Application US/08642807A
; Patent No. 6087097
;  GENERAL INFORMATION:
;    APPLICANT:  Persing, D. H.
;    TITLE OF INVENTION:  PCR DETECTION OF BORRELIA
;    TITLE OF INVENTION:  BURGDORFERI
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Schwegman, Lundberg, Woessner & Kluth, P.A.
;      STREET:  P.O. Box 2938
;      CITY:  Minneapolis
;      STATE:  MN
;      COUNTRY:  USA
;      ZIP:  55402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/642,807A
;      FILING DATE:  03-MAY-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/241,496
;      FILING DATE:  12-MAY-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Woessner, Warren D
;      REGISTRATION NUMBER:  30,440
;      REFERENCE/DOCKET NUMBER:  150.127US1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  612-373-6900
;      TELEFAX:  612-339-3061
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1100 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
US-08-642-807A-32


  Query Match             3.4%;  Score 33.4;  DB 3;  Length 1100;
  Best Local Similarity   52.1%;  Pred. No. 3.2;
  Matches   73;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy     481 aggtttgacggtgtatctagccggaggtcaaggacaagttgtaggagggaatgtggctgg 540
           || |||    | ||| |   |       | | |||   ||||| | | |    || |  | |
Db     852 AGTTTTTTTGCTGTTTACACTAATTGTTAAAGTGGAAGTTTTTGAATCCCATNTTCCAGT 793

Qy     541 ttcgttaattgcttcgggaccggtagtgttgatggctgcttcttttgcaaacgcagttta 600
           ||  ||| | || |  | |  |  |||||   |   |||  || || |     ||| ||
Db     792 TTTTTTAGTAGCCTGAGTAGTGTCAGTGTCATTAAGTGCAACTGTTACTTCTCCAGATTT 733

Qy     601 tgataggttaccgattgaag 620
           ||| | |||   | || |||
Db     732 TGAAATGTTCTTGCTTAAAG 713


RESULT   9
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2


  Query Match             3.3%;  Score 31.8;  DB 4;  Length 4403765;
  Best Local Similarity   56.1%;  Pred. No. 52;
  Matches   60;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy      81 aagcagacgaagcaaaggccgagaccactcccaccggtggagccaccagctcagccacag 140
           | || | ||  ||     ||  |  ||| | | || |  |  ||    | | | |  ||||
Db 2441955 ACGCGGGCGGTGCCGGTGCAAATCCCAATTCAACGGACGCCGCACGAATCCCCGAGACAG 2441896

Qy     141 cctctggctcttcctccggacgtcgtccacgtggtcgtcctgcaggt 187
           || ||||||  ||  | || ||||| |  || ||||  | ||  ||
Db 2441895 CCGCTGGCTAGTCAGCTGGCCGTCGGCGCCGGGGTCAGTCCGCGCGT 2441849


RESULT  10
US-08-929-329-1
; Sequence 1, Application US/08929329
; Patent No. 6120770
;  GENERAL INFORMATION:
;    APPLICANT:  Adams, John H
;    APPLICANT:  Dalton, John P
;    APPLICANT:  Kappe, Stefan
;    TITLE OF INVENTION:  Plasmodium Proteins Useful for Preparing
;    TITLE OF INVENTION:  Vaccine Compositions
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Barnes & Thornburg
;      STREET:  11 S Meridian
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  USA
;      ZIP:  46204
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/929,329
;      FILING DATE:
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Breen, John P
;      REGISTRATION NUMBER:  38,833
;      REFERENCE/DOCKET NUMBER:  835910-28685
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317) 231-7745
;      TELEFAX:  (317) 231-7433
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5433 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Plasmodium yoelii
US-08-929-329-1


  Query Match             3.2%;  Score 31.6;  DB 3;  Length 5433;
  Best Local Similarity   52.2%;  Pred. No. 23;
  Matches   70;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy     511 aggacaagttgtaggagggaatgtggctggttcgttaattgcttcgggaccggtagtgtt 570
           | || || ||| || ||   |   |||||  | |  ||      | | |  || |  |
Db    3442 AAGAAAATTTGAAGAAGCACAAAAGGCTGCGTGGGCAAAAAAAGCAGAAGAGGAAAGGAA 3501

Qy     571 gatggctgcttctttttgcaaacgcagtttatgataggttaccgattgaagaggaagaaac 630
           | |||||   || |   ||| ||||   | || ||   ||  |||||||  ||| | |
Db    3502 AAAGGCTGAAGCTGTAAAAAAAGCAGAGGAGGAAAGAAAACGAATTGAAGCTGAAAAGAA 3561

Qy     631 cccaccgccgagaa 644
           ||  |  |  ||
Db    3562 AGCAGAGGAGGAAA 3575
```

```
RESULT  11
US-08-976-259-14/c
; Sequence 14, Application US/08976259
; Patent No. 6316609
;  GENERAL INFORMATION:
;     APPLICANT:  Dillon, Patrick J.
;     APPLICANT:  Choi, Gil H.
;     APPLICANT:  Welch, Rodney A.
;     TITLE OF INVENTION:  Nucleotide Sequence of Escherichia coli
; Patent No. 6316609
;     NUMBER OF SEQUENCES:  142
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Sterne, Kessler, Goldstein & Fox P.L.L.C.
;       STREET:  1100 New York Ave, N.W., Suite 600
;       CITY:  Washington
;       STATE:  DC
;       COUNTRY:  USA
;       ZIP:  20005-3934
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;       COMPUTER:  HP Vectra 486/33
;       OPERATING SYSTEM:  MSDOS version 6.2
;       SOFTWARE:  ASCII Text
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/976,259
;       FILING DATE:  Herewith
;       CLASSIFICATION:  536
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 60/031,626 AND US 60/061,953
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Steffe, Eric K.
;       REGISTRATION NUMBER:  36,688
;       REFERENCE/DOCKET NUMBER:  1488.0740002/EKS/CBM
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (202) 371-2600
;       TELEFAX:  (202) 371-2540
;   INFORMATION FOR SEQ ID NO:  14:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  22671 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
US-08-976-259-14


  Query Match             3.2%;  Score 31.6;  DB 4;  Length 22671;
  Best Local Similarity   44.4%;  Pred. No. 47;
  Matches  127;  Conservative    0;  Mismatches  159;  Indels    0;  Gaps    0;

Qy     298 cacctacgccactcgtcgcggctgcggcgtttgcattataagcggcacgggtgcggtcac 357
           | |  | |||   |||| |||   ||  | |    ||||  || |     |||    |
Db   21116 CGCTGAAGCCGAGCGTCCCGGTATCGAGGCTGTTTTTATCTGCCGTCGCCGTGGAAGCGA 21057

Qy     358 taacgtcacgatacggcaacctgcggctccggctggtggaggtgtgattaccctgcatgg 417
           ||||  |||  ||   ||    || || | | |   |||       |     || |    ||
Db   21056 TAACCGCACCATCGAGCTGGGTGTGGTTGCCGACCGTGATATCGAAGCCGCCATCTCCGG 20997

Qy     418 tcggtttgacatttgtctttgaccggtactgcgcttccaccgcctgcaccaccgggagc 477
               |  |  |||||     | |||           | |  | |      |||||||
Db   20996 AAAACATCCCGGTTTGTTCAGCAACGGAGTCAAAGCGGCTCTTCATCTTATCCCGGGAGG 20937

Qy     478 aggaggtttgacggtgtatctagccggaggtcaaggacaagttgtaggagggaatgtggc 537
             | | | ||             ||||   ||||  ||  ||  |  || |     ||
Db   20936 CAGCGATGTAACCTGAGCCGGTCATGGAGCCAAAGGTAAAACTGCCGCCGGCAGCCACGC 20877

Qy     538 tggttcgttaattgcttcgggaccggtagtgttgatggctgcttct 583
           ||||  ||| | ||        | |  ||  || |||||||| |
Db   20876 TGGTCTGTTTACTGTCGTACTTACTGGTGTCCTGCTGGCTGCTTAT 20831


RESULT  12
US-09-199-637A-324/c
; Sequence 324, Application US/09199637A
; Patent No. 6355411
; GENERAL INFORMATION:
;  APPLICANT:  Ausubel, Frederick
;  APPLICANT:  Goodman, Howard M.
;  APPLICANT:  Rahme, Laurence G.
;  APPLICANT:  Mahajan-Miklos, Shalina
;  APPLICANT:  Tan, Man-Wah
;  APPLICANT:  Cao, Hui
;  APPLICANT:  Drenkard, Eliana
;  APPLICANT:  Tsongalis, John
;  TITLE OF INVENTION:  VIRULENCE-ASSOCIATED NUCLEIC ACID
;  TITLE OF INVENTION:  SEQUENCES AND USES THEREOF
;  FILE REFERENCE:  00786/361002
;  CURRENT APPLICATION NUMBER: US/09/199,637A
;  CURRENT FILING DATE:  1998-11-25
;  PRIOR APPLICATION NUMBER:  60/066,517
;  PRIOR FILING DATE:  1997-11-25
;  NUMBER OF SEQ ID NOS:  437
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 324
;   LENGTH: 765
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-199-637A-324


  Query Match             3.2%;  Score 31.4;  DB 4;  Length 765;
  Best Local Similarity   45.6%;  Pred. No. 10;
  Matches  110;  Conservative    0;  Mismatches  131;  Indels    0;  Gaps    0;

Qy     481 aggtttgacggtgtatctagccggaggtcaaggacaagttgtaggagggaatgtggctgg 540
           |||  || || |  |  |||  |  |   | |  ||  | |||||  |  |   || ||
Db     602 AGGCCTGGCGCTTCAGGTAGTAGTGGCCGACGTGCACTTCGTAGGCCGCCAGCAGGTTGC 543

Qy     541 ttcgttaattgcttcgggaccggtagtgttgatggctgcttcttttgcaaacgcagttta 600
             | ||     | || | |  |  ||    |   |  ||  || |||   | || | |
Db     542 GCAGGTACACCATGCGCGCCTTGGCGTCCGGGGCGTAGCGGCTGTTGGGGAAGCGGCTGG 483

Qy     601 tgataggttaccgattgaagaggaagaaaccccaccgccgagaaccaccggggtgcagca 660
           |||   |       | |  || | ||   |       |  |   | |   |  ||
Db     482 TGAGCTGGGCGAACTCGTTGAAGGAGTCGCGGGCGGCGCCCGGGTCGCGCTTGGTCATGT 423

Qy     661 gcagcagccggaggcgtctcagtcgtcggaggttacggggagtggggcccaggcgtgtga 720
           |||| || ||| |||   |||  | | | |||   ||     || |  ||||| | |||
Db     422 CCAGCGGCAGGAAGCGCGCCAGCAGGCCGCGGTCCTGGTCGAAGGAGGACAGGCCTTTGA 363

Qy     721 g 721
           |
Db     362 G 362


RESULT  13
US-09-199-637A-322/c
; Sequence 322, Application US/09199637A
; Patent No. 6355411
; GENERAL INFORMATION:
;  APPLICANT:  Ausubel, Frederick
;  APPLICANT:  Goodman, Howard M.
;  APPLICANT:  Rahme, Laurence G.
;  APPLICANT:  Mahajan-Miklos, Shalina
;  APPLICANT:  Tan, Man-Wah
;  APPLICANT:  Cao, Hui
;  APPLICANT:  Drenkard, Eliana
;  APPLICANT:  Tsongalis, John
;  TITLE OF INVENTION:  VIRULENCE-ASSOCIATED NUCLEIC ACID
;  TITLE OF INVENTION:  SEQUENCES AND USES THEREOF
;  FILE REFERENCE:  00786/361002
;  CURRENT APPLICATION NUMBER: US/09/199,637A
;  CURRENT FILING DATE:  1998-11-25
;  PRIOR APPLICATION NUMBER:  60/066,517
;  PRIOR FILING DATE:  1997-11-25
;  NUMBER OF SEQ ID NOS:  437
```

```
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 322
;   LENGTH: 1071
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-199-637A-322


  Query Match             3.2%;  Score 31.4;  DB 4;  Length 1071;
  Best Local Similarity   45.6%;  Pred. No. 12;
  Matches  110;  Conservative    0;  Mismatches  131;  Indels    0;  Gaps    0;

Qy     481 aggtttgacggtgtatctagccggaggtcaaggacaagttgtaggagggaatgtggctgg 540
           |||  || || |  |  |||  |  |   | |  ||  | |||||  |  |   || ||
Db     613 AGGCCTGGCGCTTCAGGTAGTAGTGGCCGACGTGCACTTCGTAGGCCGCCAGCAGGTTGC 554

Qy     541 ttcgttaattgcttcgggaccggtagtgttgatggctgcttcttttgcaaacgcagttta 600
             | ||     | || | |  |  ||   |   |  ||  || |||   | || | |
Db     553 GCAGGTACACCATGCGCGCCTTGGCGTCCGGGGCGTAGCGGCTGTTGGGGAAGCGGCTGG 494

Qy     601 tgataggttaccgattgaagaggaagaaaccccaccgccgagaaccaccggggtgcagca 660
           |||   |        | |  || | ||   |         |  |   | |    |  ||
Db     493 TGAGCTGGGCGAACTCGTTGAAGGAGTCGCGGGCGGCGCCCGGGTCGCGCTTGGTCATGT 434

Qy     661 gcagcagccggaggcgtctcagtcgtcggaggttacggggagtggggcccaggcgtgtga 720
           |||| || ||| |||   |||  | | | |||   ||     || |  ||||| | |||
Db     433 CCAGCGGCAGGAAGCGCGCCAGCAGGCCGCGGTCCTGGTCGAAGGAGGACAGGCCTTTGA 374

Qy     721 g 721
           |
Db     373 G 373


RESULT  14
US-09-199-637A-320
; Sequence 320, Application US/09199637A
; Patent No. 6355411
; GENERAL INFORMATION:
;  APPLICANT:  Ausubel, Frederick
;  APPLICANT:  Goodman, Howard M.
;  APPLICANT:  Rahme, Laurence G.
;  APPLICANT:  Mahajan-Miklos, Shalina
;  APPLICANT:  Tan, Man-Wah
;  APPLICANT:  Cao, Hui
;  APPLICANT:  Drenkard, Eliana
;  APPLICANT:  Tsongalis, John
;  TITLE OF INVENTION:  VIRULENCE-ASSOCIATED NUCLEIC ACID
;  TITLE OF INVENTION:  SEQUENCES AND USES THEREOF
;  FILE REFERENCE:  00786/361002
;  CURRENT APPLICATION NUMBER: US/09/199,637A
;  CURRENT FILING DATE:  1998-11-25
;  PRIOR APPLICATION NUMBER:  60/066,517
;  PRIOR FILING DATE:  1997-11-25
;  NUMBER OF SEQ ID NOS:  437
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 320
;   LENGTH: 1590
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-199-637A-320


  Query Match             3.2%;  Score 31.4;  DB 4;  Length 1590;
  Best Local Similarity   45.6%;  Pred. No. 15;
  Matches  110;  Conservative    0;  Mismatches  131;  Indels    0;  Gaps    0;

Qy     481 aggtttgacggtgtatctagccggaggtcaaggacaagttgtaggagggaatgtggctgg 540
           |||  || || |  |  |||  |  |   | |  ||  | |||||  |  |   || ||
Db     972 aggcctggcgcttcaggtagtagtggccgacgtgcacttcgtaggccgccagcaggttgc 1031

Qy     541 ttcgttaattgcttcgggaccggtagtgttgatggctgcttcttttgcaaacgcagttta 600
             | ||     | || | |  |  ||   |   |  ||  || |||   | || | |
Db    1032 gcaggtacaccatgcgcgccttggcgtccggggcgtagcggctgttggggaagcggctgg 1091

Qy     601 tgataggttaccgattgaagaggaagaaaccccaccgccgagaaccaccggggtgcagca 660
           |||   |        | |  || | ||   |         |  |   | |    |  ||
Db    1092 tgagctgggcgaactcgttgaaggagtcgcgggcggcgcccgggtcgcgcttggtcatgt 1151

Qy     661 gcagcagccggaggcgtctcagtcgtcggaggttacggggagtggggcccaggcgtgtga 720
           |||| || ||| |||   |||  | | | |||   ||     || |  ||||| | |||
Db    1152 ccagcggcaggaagcgcgccagcaggccgcggtcctggtcgaaggaggacaggcctttga 1211

Qy     721 g 721
           |
Db    1212 g 1212


RESULT  15
US-09-199-637A-1/c
; Sequence 1, Application US/09199637A
; Patent No. 6355411
; GENERAL INFORMATION:
;  APPLICANT:  Ausubel, Frederick
;  APPLICANT:  Goodman, Howard M.
;  APPLICANT:  Rahme, Laurence G.
;  APPLICANT:  Mahajan-Miklos, Shalina
;  APPLICANT:  Tan, Man-Wah
;  APPLICANT:  Cao, Hui
;  APPLICANT:  Drenkard, Eliana
;  APPLICANT:  Tsongalis, John
;  TITLE OF INVENTION:  VIRULENCE-ASSOCIATED NUCLEIC ACID
;  TITLE OF INVENTION:  SEQUENCES AND USES THEREOF
;  FILE REFERENCE:  00786/361002
;  CURRENT APPLICATION NUMBER: US/09/199,637A
;  CURRENT FILING DATE:  1998-11-25
;  PRIOR APPLICATION NUMBER:  60/066,517
;  PRIOR FILING DATE:  1997-11-25
;  NUMBER OF SEQ ID NOS:  437
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 42235
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
;   FEATURE:
;   NAME/KEY: variation
;   LOCATION: (1)...(42235)
;   OTHER INFORMATION: N is any nucleic acid.
US-09-199-637A-1


  Query Match             3.2%;  Score 31.4;  DB 4;  Length 42235;
  Best Local Similarity   45.6%;  Pred. No. 71;
  Matches  110;  Conservative    0;  Mismatches  131;  Indels    0;  Gaps    0;

Qy     481 aggtttgacggtgtatctagccggaggtcaaggacaagttgtaggagggaatgtggctgg 540
           |||  || || |  |  |||  |  |   | |  ||  | |||||  |  |   || ||
Db   28730 AGGCCTGGCGCTTCAGGTAGTAGTGGCCGACGTGCACTTCGTAGGCCGCCAGCAGGTTGC 28671

Qy     541 ttcgttaattgcttcgggaccggtagtgttgatggctgcttcttttgcaaacgcagttta 600
             | ||     | || | |  |  ||   |   |  ||  || |||   | || | |
Db   28670 GCAGGTACACCATGCGCGCCTTGGCGTCCGGGGCGTAGCGGCTGTTGGGGAAGCGGCTGG 28611

Qy     601 tgataggttaccgattgaagaggaagaaaccccaccgccgagaaccaccggggtgcagca 660
           |||   |        | |  || | ||   |         |  |   | |    |  ||
Db   28610 TGAGCTGGGCGAACTCGTTGAAGGAGTCGCGGGCGGCGCCCGGGTCGCGCTTGGTCATGT 28551

Qy     661 gcagcagccggaggcgtctcagtcgtcggaggttacggggagtggggcccaggcgtgtga 720
           |||| || ||| |||   |||  | | | |||   ||     || |  ||||| | |||
Db   28550 CCAGCGGCAGGAAGCGCGCCAGCAGGCCGCGGTCCTGGTCGAAGGAGGACAGGCCTTTGA 28491

Qy     721 g 721
           |
Db   28490 G 28490
```

Search completed: July 10, 2002, 21:34:24
Job time: 5707 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2002, 19:01:07 ; Search time 1926.27 Seconds
(without alignments)
6824.606 Million cell updates/sec

Title: US-09-823-676-1
Perfect score: 974
Sequence: 1 ccccccgacctgcctctaca.....cgaaatgcgaatattaggtt 974

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	208.6	21.4	759	10	BG581882	BG581882 EST483618
	2	206	21.1	597	9	AW585780	AW585780 EST317403
	3	198	20.3	736	10	BG583651	BG583651 EST485403
c	4	196.6	20.2	876	12	BH550632	BH550632 BOGUC18TR
	5	192.4	19.8	688	9	AW774484	AW774484 EST333635
	6	192.4	19.8	690	9	AW574000	AW574000 EST316591
	7	192.4	19.8	765	10	BG647144	BG647144 EST508763
	8	190.8	19.6	671	10	BG587531	BG587531 EST489303
	9	187.6	19.3	449	9	BE203784	BE203784 EST396460
c	10	180.4	18.5	782	12	BH419879	BH419879 BOGGX69TF
	11	179.2	18.4	534	10	BG134451	BG134451 EST467343
	12	178.4	18.3	765	12	BH481983	BH481983 BOGYI73TR
	13	177.6	18.2	488	9	AL366947	AL366947 MtBA11B10
	14	174	17.9	582	9	AW776082	AW776082 EST335147
c	15	173.4	17.8	747	12	BH566718	BH566718 BOHCV23TR
	16	166.8	17.1	745	12	BH596283	BH596283 BOGBL42TR
	17	165.4	17.0	618	9	AW980581	AW980581 EST391734
	18	163.8	16.8	715	10	BM110212	BM110212 EST557748
	19	158.8	16.3	531	9	AW560824	AW560824 EST315872
	20	158.8	16.3	555	10	BI321563	BI321563 saf12g07.
	21	158.6	16.3	395	10	BF067277	BF067277 st37f11.y
	22	158.2	16.2	652	10	BE999313	BE999313 EST431036
c	23	157.8	16.2	762	12	BH459056	BH459056 BOGKX90TR
	24	157.6	16.2	750	10	BG583687	BG583687 EST485440
	25	156.8	16.1	591	10	BF254863	BF254863 HVSMEf000
c	26	150	15.4	783	9	AW349284	AW349284 GM210004B
c	27	148.8	15.3	451	12	BH419964	BH419964 BOGQI28TF
	28	144.2	14.8	583	10	BI701170	BI701170 sag55e11.
c	29	143.8	14.8	792	12	BH513264	BH513264 BOHQC93TR
c	30	140	14.4	739	9	AW350603	AW350603 GM210008B
	31	139.8	14.4	394	9	AW774872	AW774872 EST334023
	32	138.4	14.2	628	9	AW720668	AW720668 LjNEST14h
	33	138.2	14.2	449	10	BF004270	BF004270 EST432768
c	34	137.2	14.1	814	12	BH550963	BH550963 BOHQE85TR
	35	137	14.1	542	10	BE998262	BE998262 EST429985
	36	136.4	14.0	417	10	BI419871	BI419871 LjNEST17a
	37	129	13.2	737	10	BG583795	BG583795 EST485550
c	38	126.2	13.0	309	12	BH173434	BH173434 0104B13 m
c	39	125.8	12.9	638	9	AV544413	AV544413 AV544413
	40	125.6	12.9	521	10	BG881387	BG881387 sae81e03.
	41	125.2	12.9	456	10	T43108	T43108 6371 Lambda
	42	123.2	12.6	448	9	AV422634	AV422634 AV422634
c	43	117.4	12.1	673	12	BH430946	BH430946 BOHSM94TF
	44	116.2	11.9	458	9	AI494847	AI494847 sb06b09.y
	45	115.2	11.8	335	9	BE020549	BE020549 sm45a11.y

ALIGNMENTS

RESULT 1
BG581882
LOCUS BG581882 759 bp mRNA linear EST 11-APR-2001
DEFINITION EST483618 GVN Medicago truncatula cDNA clone pGVN-66K1 5' end, mRNA sequence.
ACCESSION BG581882
VERSION BG581882.1 GI:13596946
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
REFERENCE 1 (bases 1 to 759)
AUTHORS Fedorova,M., Pierson,B.L., Samac,D.A., Vance,C.P., Gantt,G.S., Town ,C.D., Van Aken,S., Utterback,T., Cho,J. and Fraser,C.M.
TITLE ESTs from one month old nitrogen-fixing root nodules of Medicago truncatula, 2001
JOURNAL Unpublished (2001)
COMMENT Contact: Carroll P. Vance
Department of Agronomy and Plant Genetics
University of Minnesota
411 Borlaug Hall, 1991 Upper Buford Circle, St.Paul, MN 55108 USA
Tel: 612 625 5715
Fax: 651-649-5058
Email: vance004@maroon.tc.umn.edu
University of Minnesota name: M382944e TIGR sequence name: MTCDO61TK More information is available at: http://www.medicago.org
Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
FEATURES Location/Qualifiers
source 1. .759
/organism="Medicago truncatula"
/cultivar="genotype A17"
/db_xref="taxon:3880"
/clone="pGVN-66K1"
/clone_lib="GVN"
/tissue_type="N2-fixing root nodules"
/dev_stage="effective root nodules harvested one month post inoculation with Sinorhizobium meliloti"

/lab_host="E. coli strain XLOLR"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from polyA+ enriched RNA from effective root nodules harvested one month post inoculation with Sinorhizobium meliloti. The cDNA was directionally ligated into the Uni-ZAP XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-ZAP phage using Ex-Assist helper phage and propagated in XLOLR cells."

BASE COUNT 169 a 169 c 225 g 196 t
ORIGIN

Query Match 21.4%; Score 208.6; DB 10; Length 759;
Best Local Similarity 64.6%; Pred. No. 1.1e-46;
Matches 329; Conservative 0; Mismatches 174; Indels 6; Gaps 1;

```
Qy    116 ggtggagccaccagctcagccacagcctctggctcttcctccggacgtcgtccacgtggt 175
          |||  ||      | |||||  | |   | ||   |      || || |||||||| ||
Db     84 GGTTTAGAGCTTGGTTCAGCAGCCGGACCAGGAGATGTGGTAGGCCGCCGTCCACGCGGC 143

Qy    176 cgtcctgcaggttccaaaaacaaacccaaacctccgacgattataactagagatagtcct 235
          | ||  | || || |||||||  |||||| |||  ||| || ||| |||| ||| |
Db    144 AGACCGCCGGGATCGAAAAACAAGGCCAAACCACCGGTGATCATCACTCGAGAGAGTGCC 203

Qy    236 aacgtccttagatcacacgttcttgaagtcacctccggttcggacatatccgaggcagtc 295
          |||   || ||| | ||  ||||||||||  ||  |||||| ||  | |  ||    ||
Db    204 AACACTCTCAGAGCTCATATTCTTGAAGTTGCCGGCGGTTCAGATGTTTTTGAATGCGTA 263

Qy    296 tccacctacgccactcgtcgcggctgcggcgtttgcattataagcggcacgggtgcggtc 355
          || ||||| ||    || ||     | ||  | ||  || | |||||||  ||  | ||
Db    264 TCTACCTATGCTCGCCGCCGTCAACGTGGGATCTGTGTTCTCAGCGGCAGTGGCACTGTG 323

Qy    356 actaacgtcacgatacggcaacctgcggctccggctggtggaggtgtgattaccctgcat 415
          || || ||||  || |||||||| |      |||| | ||| |||||  |||| || |||
Db    324 ACAAATGTCAGTATCCGGCAACCAG------CGGCGGCTGGTGGTGTTGTTACGCTACAT 377

Qy    416 ggtcggtttgacattttgtctttgaccggtactgcgcttccaccgcctgcaccaccggga 475
          ||  | ||||| || || || ||| | ||  ||    || |||||  || || || ||| |||
Db    378 GGAAGATTTGAGATATTATCCTTGTCTGGATCTTTTCTGCCACCACCAGCTCCTCCCGGA 437

Qy    476 gcaggaggtttgacggtgtatctagccggaggtcaaggacaagttgtaggagggaatgtg 535
          |||    | || ||| | ||||| || ||||| |||||||| || || ||||| |  |||
Db    438 GCAACTAGCTTAACGATTTATCTTGCTGGAGGACAAGGACAGGTGGTGGGAGGTAGCGTG 497

Qy    536 gctggttcgttaattgcttcgggaccggtagtgttgatggctgcttcttttgcaaacgca 595
          | ||||  | | || ||  | || |||||  |  |||| || |||||||| | || |
Db    498 GTTGGTGAGCTTATCGCGGCCGGGCCGGTGATTGTGATTGCAGCTTCTTTTACTAATGTT 557

Qy    596 gtttatgataggttaccgattgaagagga 624
          | |||||| |  || ||| | ||||||||
Db    558 GCTTATGAGAAATTGCCGTTGGAAGAGGA 586
```

RESULT 2
AW585780
LOCUS AW585780 597 bp mRNA linear EST 07-SEP-2000
DEFINITION EST317403 MHAM Medicago truncatula/Glomus versiforme mixed EST library cDNA clone pMHAM-39B13, mRNA sequence.
ACCESSION AW585780
VERSION AW585780.1 GI:7265294
KEYWORDS EST.
SOURCE Medicago truncatula/Glomus versiforme mixed EST library.
ORGANISM Medicago truncatula/Glomus versiforme mixed EST library
Eukaryota; mixed EST libraries.
REFERENCE 1 (bases 1 to 597)
AUTHORS Harrison,M.J., Liu,J., Peng,H., Gonzales,M., Ellis,L., Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and Fraser,C.M.
TITLE ESTs from roots of Medicago truncatula after colonization with Glomus versiforme
JOURNAL Unpublished (2000)
COMMENT Contact: Maria J. Harrison

Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73401, USA
Tel: 580-223-5810
Fax: 580-221-7380
Email: mjharrison@noble.org
Noble EST name:N254708e
TIGR sequence name:MTDAG07TK
More information is available at. .
'http://chrysie.tamu.edu/medicago'
Seq primer: SKmod (CTA gAA CTA gtg gAT CC).

FEATURES Location/Qualifiers
source 1. .597
/organism="Medicago truncatula/Glomus versiforme mixed EST library"
/cultivar="Medicago truncatula genotype A17"
/db_xref="taxon:119092"
/clone="pMHAM-39B13"
/clone_lib="MHAM"
/tissue_type="roots colonized with Glomus versiforme"
/dev_stage="Roots harvested at 10, 17, 22, 31 and 38 days post-inoculation with Glomus versiforme. The library was made from a mixture of RNA from each of these stages."
/lab_host="E. coli strain XLOLR"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from polyA+ enriched RNA from roots harvested at 10, 17, 22, 31 and 38 days post-inoculation with Glomus versiforme. The cDNA was directionally ligated into the Unizap XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-Zap phage using Ex-assist helper phage and propagated in XLOLR cells."

BASE COUNT 159 a 136 c 149 g 153 t
ORIGIN

Query Match 21.1%; Score 206; DB 9; Length 597;
Best Local Similarity 66.7%; Pred. No. 5.1e-46;
Matches 312; Conservative 0; Mismatches 150; Indels 6; Gaps 1;

```
Qy    160 acgtcgtccacgtggtcgtcctgcaggttccaaaaacaaacccaaacctccgacgattat 219
          | |  | ||| | ||  | || || || ||||||||||||||| ||||| ||||  || ||
Db    111 AAGAAGACCAAGAGGAAGACCAGCTGGATCCAAAAACAAACCAAAACCACCGATTATCAT 170

Qy    220 aactagagatagtcctaacgtccttagatcacacgttcttgaagtcacctccggttcgga 279
          ||  | || ||  | |||| | | | | | |||||  | |||||  |   |||||  ||
Db    171 CACACGTGACAGCGCAAACGCCTTGAAAACTCACGTGATGGAAGTTGCAGACGGTTGTGA 230

Qy    280 catatccgaggcagtctccacctacgccactcgtcgcggctgcggcgtttgcattataag 339
          | |    ||    |||  || || ||| |  || ||     | ||||||||||| || ||
Db    231 CGTTGTTGAAAGTGTCAACAACTTCGCTAGACGCCGTCAAAGAGGCGTTTGCATCATGAG 290

Qy    340 cggcacgggtgcggtcactaacgtcacgatacggcaacctgcggctccggctggtggagg 399
          || || ||  |||| || ||||| ||  || |||||||  ||  ||||      |||||
Db    291 TGGGACAGGGACGGTTACAAACGTGACACTAAGGCAACCGGCTTCTCC------TGGAGC 344

Qy    400 tgtgattaccctgcatggtcggtttgacattttgtctttgaccggtactgcgcttccacc 459
          |||  |||| || |||||  |||| || ||| | ||  ||  | ||  |    ||||||||
Db    345 TGTTGTTACACTTCATGGAAGGTTCGAGATTCTTTCGCTGGCTGGATCGTTTCTTCCACC 404

Qy    460 gcctgcaccaccgggagcaggaggtttgacggtgtatctagccggaggtcaaggacaagt 519
          |||||  || || |  ||     |||||||||  | ||| |||| || || ||||| || ||
Db    405 ACCTGCTCCGCCCGCTGCTTCTGGTTTGACCATATATTTAGCGGGCGGACAAGGGCAGGT 464

Qy    520 tgtaggagggaatgtggctggttcgttaattgcttcgggaccggtagtgttgatggctgc 579
          ||| || || | ||| | |||| | || ||||||||| ||||| || ||  | ||| ||||
Db    465 TGTTGGTGGAAGTGTTGTTGGTGCTTTGATTGCTTCTGGACCTGTTGTTATTATGTCTGC 524

Qy    580 ttcttttgcaaacgcagtttatgataggttaccgattgaagaggaaga 627
          |||||||   ||||| | |||||| ||  | ||  | ||||| || ||
Db    525 TTCTTTTAGTAACGCTGCTTATGAGAGACTTCCTTTGGAAGATGATGA 572
```

```
RESULT   3
BG583651
LOCUS       BG583651                736 bp    mRNA    linear   EST 11-APR-2001
DEFINITION  EST485403 GVN Medicago truncatula cDNA clone pGVN-74E21 5' end,
            mRNA sequence.
ACCESSION   BG583651
VERSION     BG583651.1  GI:13598715
KEYWORDS    EST.
SOURCE      barrel medic.
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 736)
  AUTHORS   Fedorova,M., Pierson,B.L., Samac,D.A., Vance,C.P., Gantt,G.S., Town
            ,C.D., Van Aken,S., Utterback,T., Cho,J. and Fraser,C.M.
  TITLE     ESTs from one month old nitrogen-fixing root nodules of Medicago
            truncatula, 2001
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Carroll P. Vance
            Department of Agronomy and Plant Genetics
            University of Minnesota
            411 Borlaug Hall, 1991 Upper Buford Circle, St.Paul, MN 55108 USA
            Tel: 612 625 5715
            Fax: 651-649-5058
            Email: vance004@maroon.tc.umn.edu
            University of Minnesota name: M384729e TIGR sequence name:
            MTCEM35TK More information is available at: http://www.medicago.org
            Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
FEATURES             Location/Qualifiers
     source          1. .736
                     /organism="Medicago truncatula"
                     /cultivar="genotype A17"
                     /db_xref="taxon:3880"
                     /clone="pGVN-74E21"
                     /clone_lib="GVN"
                     /tissue_type="N2-fixing root nodules"
                     /dev_stage="effective root nodules harvested one month
                     post inoculation with Sinorhizobium meliloti"
                     /lab_host="E. coli strain XLOLR"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; cDNA was prepared from polyA+ enriched RNA from
                     effective root nodules harvested one month post
                     inoculation with Sinorhizobium meliloti. The cDNA was
                     directionally ligated into the Uni-ZAP XR vector from
                     Stratagene and packaged using Gigapack III Gold packaging
                     extracts. Plasmids containing cDNA inserts were excised
                     from the recombinant lambda-ZAP phage using Ex-Assist
                     helper phage and propagated in XLOLR cells."
BASE COUNT      157 a    168 c    219 g    192 t
ORIGIN
```

```
  Query Match             20.3%;  Score 198;  DB 10;  Length 736;
  Best Local Similarity   64.5%;  Pred. No. 9.1e-44;
  Matches  329;  Conservative    0;  Mismatches  175;  Indels    6;  Gaps    2;

Qy    116 ggtggagccaccagctcagccacagcctctggctcttcctccggacgtcgtccacgtggt 175
          |||  ||      | |||||  | |   | ||   |      || || |||||||| ||
Db     66 GGTTTAGAGCTTGGTTCAGCAGCCGGACCAGGAGATGTGGTAGGCCGCCGTCCACGCGGC 125

Qy    176 cgtcctgcaggttccaaaaacaaacccaaacctccgacgattataactagagatagtcct 235
          | ||  | || || ||||||||  |||||||| |||  ||| || ||| |||| ||| |
Db    126 AGACCGCCGGGATCGAAAAACAAGGCCAAACCACCGGTGATCATCACTCGAGAGAGTGCC 185

Qy    236 aacgtccttagatcacacgttcttgaagtcacctccggttcggacatatccgaggcagtc 295
          |||   || ||| | ||  ||||||||||  ||  |||||| ||  | |  ||      ||
Db    186 AACACTCTCAGAGCTCATATTCTTGAAGTTGCCGGCGGTTCAGATGTTTTTGAATGCGTA 245

Qy    296 tccacctacgccactcgtcgcggctgcggcgtttgcattataagcggcacgggtgcggtc 355
          || ||||| ||    || ||     | ||  | ||  || | |||||||  ||  | ||
```

```
Db    246 TCTACCTATGCTCGCCGCCGTCAACGTGGGATCTGTGTTCTCAGCGGCAGTGGCACTGTG 305

Qy    356 actaacgtcacgatacggcaacctgcggctccggctggtggaggtgtgattaccctgcat 415
          || || ||||  || |||||||| |||||  | | ||      |||||  |||| || |||
Db    306 ACAAATGTCAGTATCCGGCAACCAGCGGCGGCTGGTG-----GGTGTTGTTACGCTACAT 360

Qy    416 ggtcggtttgacattttgtctttgaccggtactgcgcttccaccgcctgcaccaccgggga 475
          ||  | ||||| || || || ||| | ||  ||   || ||||| || || || || |||
Db    361 GGAAGATTTGAGATATTATCCTTGTCTGGATCTTTTCTGCCACCACCAGCTCCTCCCGGA 420

Qy    476 gcaggaggtttgacggtgtatctagccggaggtcaaggacaagttgtaggaggga-atgt 534
          |||    | || ||| | ||||| || ||||| |||||||| || || ||||| |   ||
Db    421 GCAACTAGCTTAACGATTTATCTTGCTGGAGGACAAGGACAGGTGGTGGGAGGTAGCCGT 480

Qy    535 ggctggttcgttaattgcttcgggaccggtagtgttgatggctgcttcttttgcaaacgc 594
          || ||||  | | || ||  | || |||||  |  |||| || |||||||| | || |
Db    481 GGTTGGTGAGCTTATCGCGGCCGGGCCGGTGATTGTGATTGCAGCTTCTTTTACTAATGT 540

Qy    595 agtttatgataggttaccgattgaagagga 624
          | |||||| |  || ||| | ||||||||
Db    541 TGCTTATGAGAAATTGCCGTTGGAAGAGGA 570
```

```
RESULT   4
BH550632/c
LOCUS       BH550632                876 bp    DNA     linear   GSS 14-DEC-2001
DEFINITION  BOGUC18TR BOGU Brassica oleracea genomic clone BOGUC18, DNA
            sequence.
ACCESSION   BH550632
VERSION     BH550632.1  GI:17802412
KEYWORDS    GSS.
SOURCE      Brassica oleracea.
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 876)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: BOGUC18TF
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .876
                     /organism="Brassica oleracea"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOGUC18"
                     /clone_lib="BOGU"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT      214 a    236 c    213 g    213 t
ORIGIN
```

```
  Query Match             20.2%;  Score 196.6;  DB 12;  Length 876;
  Best Local Similarity   63.9%;  Pred. No. 2.4e-43;
  Matches  329;  Conservative    0;  Mismatches  159;  Indels   27;  Gaps    1;

Qy    141 cctctggctcttcctccggacgtcgtccacgtggtcgtcctgcaggttccaaaaacaaac 200
          |||| |||||    |||||    |||||||||| |||||| |||| || || ||||| |
Db    751 CCTCGGGCTCACTTTCCGGGAAGCGTCCACGTGGACGTCCTCCAGGGTCTAAGAACAAGC 692

Qy    201 ccaaacctccgacgattataactagagatagtcctaacgtccttagatcacacgttcttg 260
          | || || |||  |||| | || |||||||| || ||||| |||||||| || |||||||
```

```
Db    691 CGAAGCCACCGGTGATTGTGACAAGAGATAGCCCCAACGTGCTTAGATCTCATGTTCTTG 632

Qy    261 aagtcacctccggttcggacatatccgaggcagtctccacctacgccactcgtcgcggct 320
          |||||  | ||  ||  | ||||||   |||   |||  ||  |||||    ||  | ||
Db    631 AAGTCTCATCTGGAGCCGACATAATTGAGAGCGTCAACAATTACGCTCGCCGAAGAGGGA 572

Qy    321 gcggcgtttgcattataagcggcacgggtgcggtcactaacgtcacgatacggcaacctg 380
          | ||  ||  | ||||  | ||  ||  |  ||  |||||  |||||  ||||  |  ||  ||  |  |
Db    571 GAGGTGTCTCCATTCTCAGTGGTAACGGCGCGGTGGCTAACCTCACTCTCCGACAGCGGG 512

Qy    381 cggctccggctgg-------------------------tggaggtgtgattaccctgc 413
          | |  |   |||                         ||||||  |  |  ||  | |
Db    511 TGACGACTCATGGGAACAATGGTGGAACTGAAGCCGGAGCTGGAGGGATTGTGACTTTAC 452

Qy    414 atggtcggtttgacattttgtctttgaccggtactgcgcttccaccgcctgcaccaccgg 473
          ||||  |||||||  |||  |  ||  | ||  |||||  | ||||||  |||||  ||  ||  ||||
Db    451 ATGGAAGGTTTGAGATTCTTTCCATCACTGGTACGGTGCTTCCGCCGCCCGCGCCGCCGG 392

Qy    474 gagcaggaggtttgacggtgtatctagccggaggtcaaggacaagttgtaggagggaatg 533
          ||   ||  |||||  |  |  |  |  ||  ||  |  |||||  |  |  |  |||||  |  |
Db    391 GATGCGGTGGTTTATCTATCTTTGTTGCTGGTGAACAAGGTCGGATGATCGGAGGAAGAG 332

Qy    534 tggctggttcgttaattgcttcgggaccggtagtgttgatggctgcttcttttgcaaacg 593
          |||  |  |  ||  |  |||||  ||  ||  ||  |  ||||||||||  ||  ||  |  ||||
Db    331 TGGTGGCTCCGCTTGTGGCTTCTGGTCCCGTGATACTGATGGCTGCATCGTTCTCCAACG 272

Qy    594 cagtttatgataggttaccgattgaagaggaagaa 628
          ||  ||  |||  |||  |  |||  |||||||  ||||  |
Db    271 CAACTTTTGAAAGGCTTCCGCTTGAAGATGAAGGA 237
```

RESULT 5
AW774484

LOCUS	AW774484 688 bp mRNA linear EST 07-SEP-2000
DEFINITION	EST333635 KV3 Medicago truncatula cDNA clone pKV3-22P11, mRNA sequence.
ACCESSION	AW774484
VERSION	AW774484.1 GI:7718401
KEYWORDS	EST.
SOURCE	barrel medic.
ORGANISM	Medicago truncatula Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
REFERENCE	1 (bases 1 to 688)
AUTHORS	VandenBosch,K., Hurt,J., Moore,J., Beremand,P., Peng,H., Ellis,L., Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and Fraser,C.M.
TITLE	ESTs from roots of Medicago truncatula after Rhizobium inoculation
JOURNAL	Unpublished (1999)
COMMENT	Contact: VandenBosch K Department of Biology Texas A&M University College Station, TX 77843-3258, USA Tel: 409 845 7707 Fax: 409 845 2891 Email: kate@mail.bio.tamu.edu Texas A&M EST name:T257903e TIGR sequence name:MTEBA90TK More information is available at. . http://chrysie.tamu.edu/medicago Seq primer: SKmod (CTA gAA CTA gtg gAT CC).

```
FEATURES             Location/Qualifiers
     source          1. .688
                     /organism="Medicago truncatula"
                     /cultivar="genotype A17"
                     /db_xref="taxon:3880"
                     /clone="pKV3-22P11"
                     /clone_lib="KV3"
                     /tissue_type="Seedling roots"
                     /dev_stage="3 days post-inoculation with Sinorhizobium
                     meliloti"
                     /lab_host="E. coli strain XLOLR"
                     /note="Vector: pBluescript SK -; Site_1: EcoRI; Site_2:
                     XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
                     was directionally ligated into the Unizap XR vector from
                     Stratagene and packaged using Gigapack III Gold packaging
                     extracts. Plasmids containing cDNA inserts were excised
                     from the recombinant lambda-zap phage using Ex-assist
                     helper phage and propagated in XLOLR cells."
BASE COUNT      177 a    145 c    182 g    184 t
ORIGIN

  Query Match             19.8%;  Score 192.4;  DB 9;  Length 688;
  Best Local Similarity   64.2%;  Pred. No. 3.1e-42;
  Matches  308;  Conservative    0;  Mismatches  166;  Indels    6;  Gaps    1;

Qy    158 ggacgtcgtccacgtggtcgtcctgcaggttccaaaaacaaacccaaacctccgacgatt 217
          ||  ||  |  ||  |  ||  |  |||  |  ||  ||  ||  ||||||||  |||||  |||  |||
Db    142 GGTCGCAGGCCGAGGGGAAGACCTCCGGGCTCAAAGAACAAACCTAAACCACCGGTGATC 201

Qy    218 ataactagagatagtcctaacgtccttagatcacacgttcttgaagtcacctccggttcg 277
          ||  ||  |||||  |||  |  |||    |||||  |  |||  |  ||||||||  |      ||||
Db    202 ATCACAAGAGAGAGTGCAAACACACTTAGGGCTCACATCCTTGAAGTAAGTAGTGGTTGT 261

Qy    278 gacatatccgaggcagtctccacctacgccactcgtcgcggctgcggcgtttgcattata 337
          ||  |  |  ||  |  ||  |  ||  ||  ||  |      ||      |  ||  |  ||  |  |
Db    262 GATGTGTTTGACTCTGTGGCTACTTATGCAAGAAAGCGTCAAAGAGGGATCTGTGTCCTT 321

Qy    338 agcggcacgggtgcggtcactaacgtcacgatacggcaacctgcggctccggctggtgga 397
          ||  ||  |  ||  |  ||||||||||||  ||  |  ||  ||  ||  |      |  ||  |  |||
Db    322 AGTGGGAGTGGAACAGTCACTAACGTGACCTTGCGACAGCCAG------CTGCCGCAGGA 375

Qy    398 ggtgtgattaccctgcatggtcggtttgacattttgtctttgaccggtactgcgcttcca 457
          |||  |  ||  |  |||||  |||||||  ||  |  ||||||  ||||  |    ||  |||
Db    376 TCTGTAGTGACTTTACATGGAAGGTTTGAAATACTTTCTTTGTCCGGATCGTTTCTACCA 435

Qy    458 ccgcctgcaccaccgggagcaggaggtttgacggtgtatctagccggaggtcaaggacaa 517
          ||  ||  ||  ||  ||  ||||||    |  ||||||  ||||  ||  |  |||||  |||||  |||
Db    436 CCACCAGCTCCTCCAGGAGCTACAAGTTTGAGTGTGTTCCTTGGTGGAGGCCAAGGTCAA 495

Qy    518 gttgtaggagggaatgtggctggttcgttaattgcttcgggaccggtagtgttgatggct 577
          |||||  |||||  |||||  |  ||||  |  ||  |||||||  ||||||||  |  |  ||  |||
Db    496 GTTGTGGGAGGAAATGTTGTTGGTCCTTTGGTTGCTTCTGGACCGGTTATTGTCATCGCT 555

Qy    578 gcttcttttgcaaacgcagtttatgataggttaccgattgaagaggaagaaaccccaccg 637
          |  ||||||  |  ||  |  ||  |||||  |||||||||  |  ||  ||  ||  |||  |  |  |  |
Db    556 TCATCTTTTACGAATGTAGCGTATGAGAGGTTACCGTTGGATGAAGATGAATCTCTTCAG 615
```

RESULT 6
AW574000

LOCUS	AW574000 690 bp mRNA linear EST 07-SEP-2000
DEFINITION	EST316591 GVN Medicago truncatula cDNA clone pGVN-50F8, mRNA sequence.
ACCESSION	AW574000
VERSION	AW574000.1 GI:7238733
KEYWORDS	EST.
SOURCE	barrel medic.
ORGANISM	Medicago truncatula Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
REFERENCE	1 (bases 1 to 690)
AUTHORS	Fedorova,M., Pierson,B.L., Samac,D.A., Vance,C.P., Gantt,J.S., Peng ,H., Ellis,L., Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and Fraser,C.M.
TITLE	ESTs from one month old nitrogen-fixing root nodules of Medicago truncatula
JOURNAL	Unpublished (2000)
COMMENT	Contact: Carroll P. Vance Department of Agronomy and Plant Genetics University of Minnesota

411 Borlaug Hall, 1991 Upper Buford Circle, St.Paul, MN 55108 USA
Tel: 612 625 5715
Fax: 651-649-5058
Email: vance004@maroon.tc.umn.edu
Minnesota EST name:M252898e
TIGR sequence name:MTCAP28TK
More information is available at. .
'http://chrysie.tamu.edu/medicago'
Seq primer: SKmod (CTA gAA CTA gtg gAT CC).

FEATURES Location/Qualifiers
source 1. .690
/organism="Medicago truncatula"
/cultivar="genotype A17"
/db_xref="taxon:3880"
/clone="pGVN-50F8"
/clone_lib="GVN"
/tissue_type="N2-fixing root nodules"
/dev_stage="effective root nodules harvested one month post inoculation with Sinorhizobium meliloti"
/lab_host="E. coli strain XLOLR"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from polyA+ enriched RNA from effective root nodules harvested one month post inoculation with Sinorhizobium meliloti. The cDNA was directionally ligated into the Uni-ZAP XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-ZAP phage using Ex-Assist helper phage and propagated in XLOLR cells."

BASE COUNT 176 a 146 c 182 g 185 t 1 others
ORIGIN

Query Match 19.8%; Score 192.4; DB 9; Length 690;
Best Local Similarity 64.2%; Pred. No. 3.2e-42;
Matches 308; Conservative 0; Mismatches 166; Indels 6; Gaps 1;

```
Qy    158 ggacgtcgtccacgtggtcgtcctgcaggttccaaaaacaaacccaaacctccgacgatt 217
          || ||  | ||  | ||  | ||| | || || || |||||||| ||||| |||  |||
Db    136 GGTCGCAGGCCGAGGGGAAGACCTCCGGGCTCAAAGAACAAACCTAAACCACCGGTGATC 195

Qy    218 ataactagagatagtcctaacgtccttagatcacacgttcttgaagtcacctccggttcg 277
          || || ||||| ||| | |||   |||||  | ||| | |||||||| |     ||||
Db    196 ATCACAAGAGAGAGTGCAAACACACTTAGGGCTCACATCCTTGAAGTAAGTAGTGGTTGT 255

Qy    278 gacatatccgaggcagtctccacctacgccactcgtcgcggctgcggcgtttgcattata 337
          ||  | |  ||  | ||  | || || || |      ||     | ||  | ||  |  |
Db    256 GATGTGTTTGACTCTGTGGCTACTTATGCAAGAAAGCGTCAAAGAGGGATCTGTGTCCTT 315

Qy    338 agcggcacgggtgcggtcactaacgtcacgatacggcaacctgcggctccggctggtgga 397
          || || |  ||  | ||||||||||| ||  | || || || |      | || |  |||
Db    316 AGTGGGAGTGGAACAGTCACTAACGTGACCTTGCGACAGCCAG------CTGCCGCAGGA 369

Qy    398 ggtgtgattaccctgcatggtcggtttgacattttgtctttgaccggtactgcgcttcca 457
            |||  | ||  | |||||  ||||||| ||  | ||||| |||| |    || |||
Db    370 TCTGTAGTGACTTTACATGGAAGGTTTGAAATACTTTCTTTGTCCGGATCGTTTCTACCA 429

Qy    458 ccgcctgcaccaccgggagcaggaggtttgacggtgtatctagccggaggtcaaggacaa 517
          || || || || || |||||   | ||||||  ||||  || |  ||||| ||||| |||
Db    430 CCACCAGCTCCTCCAGGAGCTACAAGTTTGAGTGTGTTCCTTGGTGGAGGCCAAGGTCAA 489

Qy    518 gttgtaggagggaatgtggctggttcgttaattgcttcgggaccggtagtgttgatggct 577
          |||||  ||||| |||||  | ||||  | ||  |||||||  |||||||  |  | || |||
Db    490 GTTGTGGGAGGAAATGTTGTTGGTCCTTTGGTTGCTTCTGGACCGGTTATTGTCATCGCT 549

Qy    578 gcttcttttgcaaacgcagtttatgataggttaccgattgaagaggaagaaaccccaccg 637
            | |||||| | || | ||  ||||| |||||||| | || || || ||| | |  | |
Db    550 TCATCTTTTACGAATGTAGCGTATGAGAGGTTACCGTTGGATGAAGATGAATCTCTTCAG 609
```

RESULT 7
BG647144
LOCUS BG647144 765 bp mRNA linear EST 24-APR-2001
DEFINITION EST508763 HOGA Medicago truncatula cDNA clone pHOGA-15F24 5' end, mRNA sequence.
ACCESSION BG647144
VERSION BG647144.1 GI:13782256
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
REFERENCE 1 (bases 1 to 765)
AUTHORS Hahn,M.G., Ojanen-Reuhs,T., Samac,D., Town,C.D., Van Aken,S., Utterback,T., Cho,J. and Fraser,C.M.
TITLE ESTs from roots of Medicago truncatula treated with oligogalacturonides of DP 6-20
JOURNAL Unpublished (2001)
COMMENT Contact: Michael G. Hahn
Complex Carbohydrate Research Center
University of Georgia
220 Riverbend Road, Athens, GA 30602-4712, USA
Tel: 706-542-4457
Fax: 706-542-4412
Email: hahn@ccrc.uga.edu
G390750e TIGR sequence name: MTMBV36TK More information is available at: www.medicago.org
Seq primer: SKmod (CTA gAA CTA gtg gAT CC).

FEATURES Location/Qualifiers
source 1. .765
/organism="Medicago truncatula"
/cultivar="A17"
/db_xref="taxon:3880"
/clone="pHOGA-15F24"
/clone_lib="HOGA"
/tissue_type="3 day old seedling roots"
/dev_stage="24 hours after treatment in the dark at 26 C with 0.5 mg/ml oligogalacturonides (DP 6-20) in the presence of 100 ug/ml Gentamicin"
/lab_host="XLOLR"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into the Unizap XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-Zap phage using Ex-assist helper phage and propagated in SOLR cells."

BASE COUNT 192 a 161 c 191 g 221 t
ORIGIN

Query Match 19.8%; Score 192.4; DB 10; Length 765;
Best Local Similarity 64.2%; Pred. No. 3.3e-42;
Matches 308; Conservative 0; Mismatches 166; Indels 6; Gaps 1;

```
Qy    158 ggacgtcgtccacgtggtcgtcctgcaggttccaaaaacaaacccaaacctccgacgatt 217
          || ||  | ||  | ||  | ||| | || || || |||||||| ||||| |||  |||
Db    256 GGTCGCAGGCCGAGGGGAAGACCTCCGGGCTCAAAGAACAAACCTAAACCACCGGTGATC 315

Qy    218 ataactagagatagtcctaacgtccttagatcacacgttcttgaagtcacctccggttcg 277
          || || ||||| ||| | |||   |||||  | ||| | |||||||| |     ||||
Db    316 ATCACAAGAGAGAGTGCAAACACACTTAGGGCTCACATCCTTGAAGTAAGTAGTGGTTGT 375

Qy    278 gacatatccgaggcagtctccacctacgccactcgtcgcggctgcggcgtttgcattata 337
          ||  | |  ||  | ||  | || || || |      ||     | ||  | ||  |  |
Db    376 GATGTGTTTGACTCTGTGGCTACTTATGCAAGAAAGCGTCAAAGAGGGATCTGTGTCCTT 435

Qy    338 agcggcacgggtgcggtcactaacgtcacgatacggcaacctgcggctccggctggtgga 397
          || || |  ||  | ||||||||||| ||  | || || || |      | || |  |||
Db    436 AGTGGGAGTGGAACAGTCACTAACGTGACCTTGCGACAGCCAG------CTGCCGCAGGA 489

Qy    398 ggtgtgattaccctgcatggtcggtttgacattttgtctttgaccggtactgcgcttcca 457
            |||  | ||  | |||||  ||||||| ||  | ||||| |||| |    || |||
Db    490 TCTGTAGTGACTTTACATGGAAGGTTTGAAATACTTTCTTTGTCCGGATCGTTTCTACCA 549
```

```
Qy     458 ccgcctgcaccaccgggagcaggaggtttgacggtgtatctagccggaggtcaaggacaa 517
           || || || || || |||||   | ||||||  ||||  || |  ||||| ||||| |||
Db     550 CCACCAGCTCCTCCAGGAGCTACAAGTTTGAGTGTGTTCCTTGGTGGAGGCCAAGGTCAA 609

Qy     518 gttgtaggagggaatgtggctggttcgttaattgcttcgggaccggtagtgttgatggct 577
           ||||| ||||| ||||| | |||| | ||  |||||||| ||||||||  |  | || |||
Db     610 GTTGTGGGAGGAAATGTTGTTGGTCCTTTGGTTGCTTCTGGACCGGTTATTGTCATCGCT 669

Qy     578 gcttcttttgcaaacgcagtttatgataggttaccgattgaagaggaagaaaccccaccg 637
           | |||||| | || | ||  ||||| |||||||||| | || || || ||| | |  | |
Db     670 TCATCTTTTACGAATGTAGCGTATGAGAGGTTACCGTTGGATGAAGATGAATCTCTTCAG 729
```

RESULT 8
BG587531

LOCUS	BG587531 671 bp mRNA linear EST 11-APR-2001
DEFINITION	EST489303 MHAM Medicago truncatula/Glomus versiforme mixed EST library cDNA clone pMHAM-50G4 5' end, mRNA sequence.
ACCESSION	BG587531
VERSION	BG587531.1 GI:13602595
KEYWORDS	EST.
SOURCE	Medicago truncatula/Glomus versiforme mixed EST library.
ORGANISM	Medicago truncatula/Glomus versiforme mixed EST library Eukaryota; mixed EST libraries.
REFERENCE	1 (bases 1 to 671)
AUTHORS	Harrison,M.J., Liu,J., Town,C.D., Van Aken,S., Utterback,T., Cho,J. and Fraser,C.M.
TITLE	ESTs from roots of Medicago truncatula after colonization with Glomus versiforme, 2001
JOURNAL	Unpublished (2001)
COMMENT	Contact: Harrison M.J. Plant Biology Division The Samuel Roberts Noble Foundation 2510 Sam Noble Parkway, Ardmore, OK 73401 Tel: 580-223-5810 Fax: 580-221-7380 Email: mjharrison@noble.org Noble EST name: N386601e TIGR sequence name: MTDDH38TK More information is available at: http://www.medicago.org Seq primer: SKmod (CTA gAA CTA gtg gAT CC).

FEATURES	Location/Qualifiers
source	1. .671 /organism="Medicago truncatula/Glomus versiforme mixed EST library" /cultivar="Medicago truncatula genotype A17" /db_xref="taxon:119092" /clone="pMHAM-50G4" /clone_lib="MHAM" /tissue_type="roots colonized with Glomus versiforme" /dev_stage="Roots harvested at 10, 17, 22, 31 and 38 days post-inoculation with Glomus versiforme. The library was made from a mixture of RNA from each of these stages." /lab_host="E. coli strain XLOLR" /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from polyA+ enriched RNA from roots harvested at 10, 17, 22, 31 and 38 days post-inoculation with Glomus versiforme. The cDNA was directionally ligated into the Unizap XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-Zap phage using Ex-assist helper phage and propagated in XLOLR cells."

BASE COUNT 174 a 140 c 182 g 175 t
ORIGIN

Query Match 19.6%; Score 190.8; DB 10; Length 671;
Best Local Similarity 64.0%; Pred. No. 8.6e-42;
Matches 307; Conservative 0; Mismatches 167; Indels 6; Gaps 1;

```
Qy     158 ggacgtcgtccacgtggtcgtcctgcaggttccaaaaacaaacccaaacctccgacgatt 217
           || ||  | ||  | ||  | ||| | || || || |||||||| ||||| |||  |||
Db     142 GGTCGCAGGCCGAGGGGAAGACCTCCGGGCTCAAAGAACAAACCTAAACCACCGGTGATC 201

Qy     218 ataactagagatagtcctaacgtccttagatcacacgttcttgaagtcacctccggttcg 277
           || || ||||| ||| | |||   ||||||  | ||| | |||||||| |     ||||
Db     202 ATCACAAGAGAGAGTGCAAACACACTTAGGGCTCACATCCTTGAAGTAAGTAGTGGTTGT 261

Qy     278 gacatatccgaggcagtctccacctacgccactcgtcgcggctgcggcgtttgcattata 337
           ||  | |  ||  | ||  | || || || |     ||     | ||  | ||  | |
Db     262 GATGTGTTTGACTCTGTGGCTACTTATGCAAGAAAGCGTCAAAGAGGGATCTGTGTCCTT 321

Qy     338 agcggcacgggtgcggtcactaacgtcacgatacggcaacctgcggctccggctggtgga 397
           || || |  ||  | ||||||||||| ||  | || || || |      | || |  |||
Db     322 AGTGGGAGTGGAACAGTCACTAACGTGACCTTGCGACAGCCAG------CTGCCGCAGGA 375

Qy     398 ggtgtgattaccctgcatggtcggtttgacattttgtctttgaccggtactgcgcttcca 457
           |||  | ||  | |||||  |||||| ||  | |||||| ||||  |    || |||
Db     376 TCTGTAGTGACTTTACATGGAAGGTTTGAAATACTTTCTTTGTCCGGATCGTTTCTACCA 435

Qy     458 ccgcctgcaccaccgggagcaggaggtttgacggtgtatctagccggaggtcaaggacaa 517
           || || || || || |||||   | ||||||  ||||  || |  ||||| ||||| |||
Db     436 CCACCAGCTCCTCCAGGAGCTACAAGTTTGAGTGTGTTCCTTGGTGGAGGCCAAGGTCAA 495

Qy     518 gttgtaggagggaatgtggctggttcgttaattgcttcgggaccggtagtgttgatggct 577
           ||||| ||||| ||||| | |||| | ||  |||||||| ||||||||  |  | || |||
Db     496 GTTGTGGGAGGAAATGTTGTTGGTCCTTTGGTTGCTTCTGGACCGGGTATTGTCATCGCT 555

Qy     578 gcttcttttgcaaacgcagtttatgataggttaccgattgaagaggaagaaaccccaccg 637
           | |||||| | || | ||  ||||| |||||||||| | || || || ||| | |  | |
Db     556 TCATCTTTTACGAATGTAGCGTATGAGAGGTTACCGTTGGATGAAGATGAATCTCTTCAG 615
```

RESULT 9
BE203784

LOCUS	BE203784 449 bp mRNA linear EST 05-SEP-2000
DEFINITION	EST396460 KV0 Medicago truncatula cDNA clone pKV0-11B13, mRNA sequence.
ACCESSION	BE203784
VERSION	BE203784.1 GI:8747063
KEYWORDS	EST.
SOURCE	barrel medic.
ORGANISM	Medicago truncatula Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
REFERENCE	1 (bases 1 to 449)
AUTHORS	VandenBosch,K., Endre,G., Hur,J., Moore,J., Beremand,P., Ellis,L., Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and Fraser,C.M.
TITLE	ESTs from uninoculated seedling roots of Medicago truncatula
JOURNAL	Unpublished (1999)
COMMENT	Contact: VandenBosch K Department of Biology Texas A&M University College Station, TX 77843-3258, USA Tel: 409 845 7707 Fax: 409 845 2891 Email: kate@mail.bio.tamu.edu Texas A&M University name:T263556e TIGR sequence name:MTGAG07TK More information is available at. . http://chrysie.tamu.edu/medicago Seq primer: SKmod (CTA gAA CTA gtg gAT CC).

FEATURES	Location/Qualifiers
source	1. .449 /organism="Medicago truncatula" /cultivar="genotype A17" /db_xref="taxon:3880" /clone="pKV0-11B13" /clone_lib="KV0" /tissue_type="Seedling roots" /dev_stage="Immediately prior to inoculation with Sinorhizobium meliloti (0 hour)"

/lab_host="E.coli strain XLOLR"
/note="Vector: pBluescript SK -; Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into the Unizap XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-Zap phage using Ex-assist helper phage and propagated in XLOLR cells."
BASE COUNT 102 a 100 c 114 g 133 t
ORIGIN

Query Match 19.3%; Score 187.6; DB 9; Length 449;
Best Local Similarity 66.4%; Pred. No. 5.7e-41;
Matches 287; Conservative 0; Mismatches 139; Indels 6; Gaps 1;

```
Qy    196 caaacccaaacctccgacgattataactagagatagtcctaacgtccttagatcacacgt 255
          |||||| ||||| |||| || || ||  | || ||  | |||| | | | | | ||||||
Db      1 CAAACCAAAACCACCGATTATCATCACACGTGACAGCGCAAACGCCTTGAAAACTCACGT 60

Qy    256 tcttgaagtcacctccggttcggacatatccgaggcagtctccacctacgccactcgtcg 315
          | ||||| |  ||||| ||| |    ||   |||  || || ||| |  || ||
Db     61 GATGGAAGTTGCAGACGGTTGTGACGTTGTTGAAAGTGTCAACAACTTCGCTAGACGCCG 120

Qy    316 cggctgcggcgtttgcattataagcggcacgggtgcggtcactaacgtcacgatacggca 375
            | ||||||||||| || || || || ||  |||| || ||||| ||  || ||||
Db    121 TCAAAGAGGCGTTTGCATCATGAGTGGGACAGGGACGGTTACAAACGTGACACTAAGGCA 180

Qy    376 acctgcggctccggctggtggaggtgtgattaccctgcatggtcggtttgacattttgtc 435
          ||| ||  ||||      ||||| |||  |||| || |||||  |||| || ||| | ||
Db    181 ACCGGCTTCTCC------TGGAGCTGTTGTTACACTTCATGGAAGGTTCGAGATTCTTTC 234

Qy    436 tttgaccggtactgcgcttccaccgcctgcaccaccgggagcaggaggtttgacggtgta 495
            || | ||  |    |||||||| ||||| || || |  ||    ||||||||  | ||
Db    235 GCTGGCTGGATCGTTTCTTCCACCACCTGCTCCGCCCGCTGCTTCTGGTTTGACCATATA 294

Qy    496 tctagccggaggtcaaggacaagttgtaggagggaatgtggctggttcgttaattgcttc 555
          | |||| || || ||| | || ||||| || || | ||| | |||| | || ||||||||
Db    295 TTTAGCGGGCGGACAAAGGCAGGTTGTTGGTGGAAGTGTTGTTGGTGCTTTGATTGCTTC 354

Qy    556 gggaccggtagtgttgatggctgcttcttttgcaaacgcagtttatgataggttaccgat 615
           ||||| || ||  | ||| ||||||||||   ||||| | |||||| ||  | ||  |
Db    355 TGGACCTGTTGTTATTATGTCTGCTTCTTTTAGTAACGCTGCTTATGAGAGACTTCCTTT 414

Qy    616 tgaagaggaaga 627
           ||||| || ||
Db    415 GGAAGATGATGA 426
```

RESULT 10
BH419879/c
LOCUS BH419879 782 bp DNA linear GSS 12-DEC-2001
DEFINITION BOGGX69TF BOGG Brassica oleracea genomic clone BOGGX69, DNA sequence.
ACCESSION BH419879
VERSION BH419879.1 GI:17605607
KEYWORDS GSS.
SOURCE Brassica oleracea.
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 782)
AUTHORS Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished (2001)
COMMENT Other_GSSs: BOGGX69TR
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.
FEATURES Location/Qualifiers
source 1. .782
/organism="Brassica oleracea"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BOGGX69"
/clone_lib="BOGG"
/note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT 191 a 216 c 182 g 193 t
ORIGIN

Query Match 18.5%; Score 180.4; DB 12; Length 782;
Best Local Similarity 62.2%; Pred. No. 6.9e-39;
Matches 317; Conservative 0; Mismatches 166; Indels 27; Gaps 1;

```
Qy    146 ggctcttcctccggacgtcgtccacgtggtcgtcctgcaggttccaaaaacaaacccaaa 205
          || ||      ||||    ||||||||||      |      || || || |||||  | ||
Db    757 GGATCACTTCCCGGGAAGCGTCCACGTGACGTCCTCCGGGGATCGAAGAACAAGACGAAG 698

Qy    206 cctccgacgattataactagagatagtcctaacgtccttagatcacacgttcttgaagtc 265
          || |||  |||| | || ||||||||  ||||||||||||||||| || ||||||||||||
Db    697 CCACCGGTGATTGTGACAAGAGATAGCCCTAACGTCCTTAGATCTCATGTTCTTGAAGTC 638

Qy    266 acctccggttcggacatatccgaggcagtctccacctacgccactcgtcgcggctgcggc 325
           | || ||  | ||||||   |||  |||  ||| |||||    ||  | ||  | ||
Db    637 TCATCTGGAGCCGACATAGTTGAGAGCGTCAACACTTACGCTCGCCGGAGAGGGAGAGGT 578

Qy    326 gtttgcattataagcggcacgggtgcggtcactaacgtcacgatacggcaacctg----- 380
          || | |  |  | || || |   ||  ||||  ||||||||    | || || || |
Db    577 GTCTCCGTGCTCAGTGGTAACGGCACGGTGGCTAACGTCGTTCTCCGTCAGCCGGTGACG 518

Qy    381 ----------------------cggctccggctggtggaggtgtgattaccctgcatggt 418
                               | |   | |  | |||||| ||  | ||  | |||||
Db    517 ATTCATGGGAATAATGGCGGAACTGGAGCCGGAGTTGGAGGGGTTGTGACTTTACATGGA 458

Qy    419 cggtttgacattttgtctttgaccggtactgcgcttccaccgcctgcaccaccgggagca 478
           ||||||||||  | ||  | || ||||| | |||||| |||||||| || |||||| ||
Db    457 AAGTTTGACATTCTTTCCATCACTGGTACGGTGCTTCCCCCGCCTGCGCCGCCGGGATCA 398

Qy    479 ggaggtttgacggtgtatctagccggaggtcaaggacaagttgtaggagggaatgtggct 538
          || ||  || |  | | |||  | || || ||||| || ||  | || || |  ||||
Db    397 GGTGGGCTGTCAATCTTTCTTTCTGGTGGACAAGGTCAGGTCATCGGCGGAAGCGTGGTG 338

Qy    539 ggttcgttaattgcttcgggaccggtagtgttgatggctgcttcttttgcaaacgcagtt 598
          | | || |  | ||||| || || ||  |  |||||||||| || ||  | ||||||  |
Db    337 GCTCCGCTTGTGGCTTCAGGTCCAGTTATACTGATGGCTGCATCGTTCTCTAACGCAACT 278

Qy    599 tatgataggttaccgattgaagaggaagaa 628
          | ||| ||| | ||  || |||| ||||||
Db    277 TTTGAAAGGCTTCCACTTAAAGATGAAGAA 248
```

RESULT 11
BG134451
LOCUS BG134451 534 bp mRNA linear EST 31-JAN-2001
DEFINITION EST467343 tomato crown gall Lycopersicon esculentum cDNA clone cTOE16I16 5' sequence, mRNA sequence.
ACCESSION BG134451
VERSION BG134451.1 GI:12634639
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE 1 (bases 1 to 534)
AUTHORS van der Hoeven,R., Sun,H., Cho,J., Utterback,T., Hansen,C., Ronning

```
            ,C. and Tanksley,S.
  TITLE     Generation of ESTs from tomato crown gall tissue
  JOURNAL   Unpublished (2001)
COMMENT     Contact: CUGI
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html.
FEATURES             Location/Qualifiers
     source          1. .534
                     /organism="Lycopersicon esculentum"
                     /cultivar="TA496"
                     /db_xref="taxon:4081"
                     /clone="cTOE16I16"
                     /clone_lib="tomato crown gall"
                     /tissue_type="crown gall"
                     /dev_stage="crown galls from full-grown plants (8 wks old
                     )"
                     /lab_host="SOLR"
                     /note="Vector: pBluescript SK(-); Site_1: EcoR1; Site_2:
                     Xhol; Four wk old greenhouse plants were stab inoculated
                     on stem with Agrobacterium tumefaciens C58 (Dr. T.J. Burr,
                     Cornell U.). Galls were allowed to develop for another 4
                     wks, when gall tissue was frozen in  liquid nitrogen."
BASE COUNT      119 a    137 c    138 g    140 t
ORIGIN


  Query Match             18.4%;  Score 179.2;  DB 10;  Length 534;
  Best Local Similarity   62.7%;  Pred. No. 1.3e-38;
  Matches  302;  Conservative    0;  Mismatches  168;  Indels   12;  Gaps    1;

Qy    146 ggctcttcctccggacgtcgtccacgtggtcgtcctgcaggttccaaaaacaaacccaaa 205
          |||    |  ||   || ||||| || || |||||  | || || ||||| || |||||
Db     37 GGCGGAGCTCCCAATCGCCGTCCTCGAGGCCGTCCGCCTGGATCTAAAAATAAGCCCAAG 96

Qy    206 cctccgacgattataactagagatagtcctaacgtccttagatcacacgttcttgaagtc 265
          |||||  |   ||  | || |||||||  |||||||  ||  |||| ||||| ||||||||
Db     97 CCTCCAATCATCGTGACGAGAGATACGCCTAACGCACTCCGATCTCACGTGCTTGAAGTT 156

Qy    266 acctccggttcggacatatccgaggcagtctccacctacgccactcgtcgcggctgcggc 325
          |  ||| |   || ||    ||     ||||||  |||||  |   ||  | ||  | ||
Db    157 TCGACCGATGTTGATATCATGGAAAGTATCTCCAATTACGCAAGGCGGAGAGGGAGAGGT 216

Qy    326 gtttgcattataagcggcacgggtgcggtcactaacgtcacgatacggcaacctgcggct 385
          ||||| ||| | || || |  ||  | || || |||||||    | || || |||||
Db    217 GTTTGTATTCTTAGTGGTAGCGGCACAGTTACCAACGTCAACCTTCGTCAGCCTGC---- 272

Qy    386 ccggctggtggaggtgtgattaccctgcatggtcggtttgacattttgtctttgaccggt 445
              || | ||||  | || || || || || || || || ||  |      |  | |||
Db    273 --------TGCAAGTGTAGTCACACTCCACGGACGTTTCGAAATACTTAGCCTCTCAGGT 324

Qy    446 actgcgcttccaccgcctgcaccaccgggagcaggaggtttgacggtgtatctagccgga 505
          || | |||||| ||||||||||| || |   |  | ||  |  |  | | | || | ||
Db    325 ACGGTGCTTCCTCCGCCTGCACCGCCCGCCTCCAGTGGGATCTCTATATTTTTATCAGGT 384

Qy    506 ggtcaaggacaagttgtaggagggaatgtggctggttcgttaattgcttcgggaccggta 565
          || |||||||||| || |||||    || |  ||  | || || || || || || ||
Db    385 GGACAAGGACAAGTGGTTGGAGGATCCGTTGTAGGGCCTTTGATCGCATCAGGTCCAGTC 444

Qy    566 gtgttgatggctgcttcttttgcaaacgcagtttatgataggttaccgattgaagaggaa 625
          || || |||||||| |||||||| || || || | |||  |  | ||  | || || ||
Db    445 GTCTTAATGGCTGCCTCTTTTGCTAATGCTGTATTTGAACGACTTCCCTTGGAGGAAGAT 504

Qy    626 ga 627
          ||
Db    505 GA 506


RESULT  12
BH481983
LOCUS       BH481983                765 bp    DNA     linear   GSS 13-DEC-2001
DEFINITION  BOGYI73TR BOGY Brassica oleracea genomic clone BOGYI73, DNA
            sequence.
ACCESSION   BH481983
VERSION     BH481983.1  GI:17690087
KEYWORDS    GSS.
SOURCE      Brassica oleracea.
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 765)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: BOGYI73TF
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .765
                     /organism="Brassica oleracea"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOGYI73"
                     /clone_lib="BOGY"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT      181 a    161 c    215 g    208 t
ORIGIN


  Query Match             18.3%;  Score 178.4;  DB 12;  Length 765;
  Best Local Similarity   63.2%;  Pred. No. 2.5e-38;
  Matches  306;  Conservative    0;  Mismatches  151;  Indels   27;  Gaps    1;

Qy    172 tggtcgtcctgcaggttccaaaaacaaacccaaacctccgacgattataactagagatag 231
          || ||||||| |||| || || ||||| || |  || |||  |||| | || ||||||||
Db      1 TGTTCGTCCTCCAGGGTCTAAGAACAAGCCGATGCCACCGGTGATTGTGACAAGAGATAG 60

Qy    232 tcctaacgtccttagatcacacgttcttgaagtcacctccggttcggacatatccgaggc 291
          || ||||| |||||||| || |||||||||||| | || ||  | ||||||   |||
Db     61 CCCCAACGTGCTTAGATCTCATGTTCTTGAAGTCTCATCTGGAGCCGACATAATTGAGAG 120

Qy    292 agtctccacctacgccactcgtcgcggctgcggcgtttgcattataagcggcacgggtgc 351
          |||  ||  |||||    ||  | ||  | || || | |||| | || || |  || ||
Db    121 CGTCAACAATTACGCTCGCCGAAGAGGGAGAGGTGTCTCCATTCTCAGTGGTAACGGCGC 180

Qy    352 ggtcactaacgtcacgatacggcaacctgcggctccggctgg------------------ 393
          |||  |||| |||  | || || |  | | |  |   |||
Db    181 GGTGGCTAACCTCACTCTCCGACAGCGGGTGACGACTCATGGGAACAATGGTGGAACTGA 240

Qy    394 ---------tggaggtgtgattaccctgcatggtcggtttgacattttgtctttgaccgg 444
                  ||||||  |  | ||  | ||||  |||||| ||| | ||  | || ||
Db    241 AGCCGGAGCTGGAGGGATTGTGACTTTACATGGAAGGTTTGAGATTCTTTCCATCACTGG 300

Qy    445 tactgcgcttccaccgcctgcaccaccgggagcaggaggtttgacggtgtatctagccgg 504
          ||| | |||||| |||| || || ||||||   || ||||  |  | | | | || ||
Db    301 TACGGTGCTTCCGCCGCCCGCGCCGCCGGGATGCGGTGGTTTATCTATCTTTGTTGCTGG 360

Qy    505 aggtcaaggacaagttgtaggagggaatgtggctggttcgttaattgcttcgggaccggt 564
          |  ||||| |   |  | ||||| |  |||  | | || |  | ||||| || || ||
Db    361 TGAACAAGGTCGGATGATCGGAGGAAGAGTGGTGGCTCCGCTTGTGGCTTCTGGTCCCGT 420

Qy    565 agtgttgatggctgcttcttttgcaaacgcagtttatgataggttaccgattgaagagga 624
            |  |||||||||| || ||  | ||||||  || ||| ||| | ||| |||||||| ||
Db    421 GATACTGATGGCTGCATCGTTCTCCAACGCAACTTTTGAAAGGCTTCCGCTTGAAGATGA 480

Qy    625 agaa 628
```

```
          || |
Db    481 AGGA 484


RESULT  13
AL366947
LOCUS       AL366947                488 bp    mRNA    linear   EST 03-AUG-2000
DEFINITION  MtBA11B10F1 MtBA Medicago truncatula cDNA clone MtBA11B10 T3, mRNA
            sequence.
ACCESSION   AL366947
VERSION     AL366947.1  GI:9666700
KEYWORDS    EST.
SOURCE      barrel medic.
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 488)
  AUTHORS   Journet,E.P., Crespeau,H., van-Tuinen,D., Gouzy,J., Jaillon,O.,
            Niebel,A., Carreau,V., Chatagnier,O., Kahn,D., Gianinazzi-Pearson
            ,V. and Gamas,P.
  TITLE     Medicago truncatula ESTs from nitrogen-starved roots
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de
            Biologie Moleculaire des Relations Plantes-Microorganismes,
            CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email :
            Mt-est@toulouse.inra.fr Website :
            http://sequence.toulouse.inra.fr/Mtruncatula.html).
FEATURES             Location/Qualifiers
     source          1. .488
                     /organism="Medicago truncatula"
                     /cultivar="Jemalong"
                     /db_xref="taxon:3880"
                     /clone="MtBA11B10"
                     /clone_lib="MtBA"
                     /tissue_type="root tips"
                     /dev_stage="harvested after 3 days of N-starvation"
                     /note="Vector: pBluescript pSK; Site_1: EcoRI; Site_2:
                     XhoI; Plants were grown in an aeroponic chamber for 14
                     days on nitrogen-rich medium followed by 3 days on N-free
                     medium. RNA was extracted from root tips (1-3 cm). cDNA
                     was prepared from polyA+ enriched RNA. The cDNA was
                     directionally ligated into Uni-zapXR vector from
                     Stratagene and packaged using Gigapack Gold packaging
                     extracts. Plasmids containing cDNA inserts were
                     mass-excised from phage stocks using ExAssit helper phage
                     and propagated in SOLR cells. Clone ordering and
                     sequencing was performed by the Centre National de
                     Sequencage (Genoscope, Evry, France)."
BASE COUNT      126 a    116 c    136 g    110 t
ORIGIN


  Query Match             18.2%;  Score 177.6;  DB 9;  Length 488;
  Best Local Similarity   64.1%;  Pred. No. 3.4e-38;
  Matches  286;  Conservative    0;  Mismatches  154;  Indels    6;  Gaps    1;

Qy    189 ccaaaaacaaacccaaacctccgacgattataactagagatagtcctaacgtccttagat 248
          | |||||||||| ||||| || |  || || || || || ||  | ||||  ||  |||
Db     38 CGAAAAACAAACCAAAACCACCTATCATCATCACGAGGGACAGCGCGAACGCACTCCGAT 97

Qy    249 cacacgttcttgaagtcacctccggttcggacatatccgaggcagtctccacctacgcca 308
          | |||||  | |||||  |    || |  |||||    ||    ||  |   ||  ||
Db     98 CCCACGTGATGGAAGTTGCAAATGGATGTGACATCATGGAAAGTGTGACGGTCTTTGCGC 157

Qy    309 ctcgtcgcggctgcggcgtttgcattataagcggcacgggtgcggtcactaacgtcacga 368
            |  |     | || || |||||  | |||||  |  ||  | |||||  ||||| ||
Db    158 GAAGGAGGCAGCGTGGTGTCTGCATCCTTAGCGGAAGTGGGACCGTCACAAACGTGACTC 217

Qy    369 tacggcaacctgcggctccggctggtggaggtgtgattaccctgcatggtcggtttgaca 428
          | || ||||| ||  | ||       ||| |  ||  | || || |||||  | ||||| |
Db    218 TCCGTCAACCAGCATCGCC------TGGTGCGGTAGTCACACTTCATGGAAGATTTGAGA 271

Qy    429 ttttgtctttgaccggtactgcgcttccaccgcctgcaccaccgggagcaggaggtttga 488
          | || || ||  | ||  ||   || || |||||||| ||||| |  ||   ||| ||
Db    272 TATTATCATTATCTGGCTCTTTCCTGCCGCCGCCTGCTCCACCAGCGGCGTCAGGATTAG 331

Qy    489 cggtgtatctagccggaggtcaaggacaagttgtaggagggaatgtggctggttcgttaa 548
          |  | |||||||| || || |||||||| || || || || |  ||||  ||  ||||
Db    332 CCATATATCTAGCTGGTGGACAAGGACAGGTCGTCGGTGGTAGCGTGGTGGGACCGTTGT 391

Qy    549 ttgcttcgggaccggtagtgttgatggctgcttcttttgcaaacgcagtttatgataggt 608
          | ||||| || ||||| ||  | ||||| ||||| |||| ||| || | |||||| |||
Db    392 TAGCTTCCGGTCCGGTTGTTATCATGGCAGCTTCCTTTGGAAATGCTGCTTATGAAAGGC 451

Qy    609 taccgattgaagaggaagaaacccca 634
          ||||  | ||||| |||||||| |||
Db    452 TACCTTTAGAAGATGAAGAAACACCA 477


RESULT  14
AW776082
LOCUS       AW776082                582 bp    mRNA    linear   EST 07-SEP-2000
DEFINITION  EST335147 DSIL Medicago truncatula cDNA clone pDSIL-6E7, mRNA
            sequence.
ACCESSION   AW776082
VERSION     AW776082.1  GI:7765895
KEYWORDS    EST.
SOURCE      barrel medic.
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 582)
  AUTHORS   Fedorova,M., Pierson,B.L., Samac,D.A., Vance,C.P., Gantt,G.S., Peng
            ,H., Ellis,L., Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S.,
            Holt,I.E. and Fraser,C.M.
  TITLE     ESTs from leaves of Medicago truncatula after inoculation with
            Colletotrichum trifolii
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Deborah A. Samac
            Department of Plant Pathology
            University of Minnesota
            495 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108, USA
            Tel: 612 625 1243
            Fax: 651 649 5058
            Email: debbys@puccini.crl.umn.edu
            Minnesota sequence name:M259415e
            TIGR sequence name:MTFAM28TK
            More information is available at. .
            http://chrysie.tamu.edu/medicago
            Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
FEATURES             Location/Qualifiers
     source          1. .582
                     /organism="Medicago truncatula"
                     /cultivar="genotype A17"
                     /db_xref="taxon:3880"
                     /clone="pDSIL-6E7"
                     /clone_lib="DSIL"
                     /tissue_type="leaves infected with Colletotrichum
                     trifolii"
                     /dev_stage="cotyledons and primary leaves harvested 5 and
                     8 days after inoculation with Colletotrichum trifolii"
                     /lab_host="E. coli strain XLOLR"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; cDNA was prepared from polyA+ enriched RNA from
                     cotyledons and primary leaves harvested 5 and 8 days after
                     inoculation with Colletotrichum trifolii. The cDNA was
                     directionally ligated into the Uni-ZAP XR vector from
                     Stratagene and packaged using Gigapack III Gold packaging
```

extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-ZAP phage using Ex-Assist helper phage and propagated in XLOLR cells. Note: EST may be of fungal origin."
BASE COUNT 148 a 148 c 156 g 129 t 1 others
ORIGIN

Query Match 17.9%; Score 174; DB 9; Length 582;
Best Local Similarity 63.6%; Pred. No. 3.7e-37;
Matches 283; Conservative 0; Mismatches 156; Indels 6; Gaps 1;

```
Qy    116 ggtggagccaccagctcagccacagcctctggctcttcctccggacgtcgtccacgtggt 175
          |||  ||       | |||||  | |   | ||    |       || || |||||||| ||
Db    144 GGTTTAGAGCTTGGTTCAGCAGCCGGACCAGGAGATGTGGTAGGCCGCCGTCCACGCGGC 203

Qy    176 cgtcctgcaggttccaaaaacaaacccaaacctccgacgattataactagagatagtcct 235
          | ||  | || || ||||||||  ||||||| |||  ||| || ||| |||| ||| |
Db    204 AGACCGCCGGGATCGAAAAACAAGGCCAAACCACCGGTGATCATCACTCGAGAGAGTGCC 263

Qy    236 aacgtccttagatcacacgttcttgaagtcacctccggttcggacatatccgaggcagtc 295
          |||   || ||| | ||  ||||||||||  ||  ||||||  ||  | |  ||    ||
Db    264 AACACTCTCAGAGCTCATATTCTTGAAGTTGCCGGCGGTTCAGATGTTTTTGAATGCGTA 323

Qy    296 tccacctacgccactcgtcgcggctgcggcgtttgcattataagcggcacgggtgcggtc 355
          || ||||| ||     || ||     | ||  | ||  || | |||||||  ||   | ||
Db    324 TCTACCTATGCTCGCCGCCGTCAACGTGGGATCTGTGTTCTCAGCGGCAGTGGCACTGTG 383

Qy    356 actaacgtcacgatacggcaacctgcggctccggctggtggaggtgtgattaccctgcat 415
          || || ||||  || ||||||||      |  |||| | ||| |||||  |||| || |||
Db    384 ACAAATGTCAGTATCCGGCAAC------CANCGGCGGCTGGTGGTGTTGTTACGCTACAT 437

Qy    416 ggtcggtttgacattttgtctttgaccggtactgcgcttccaccgcctgcaccaccgggga 475
          ||  | ||||| || || || ||| | ||  ||   || ||||| || || || || |||
Db    438 GGAAGATTTGAGATATTATCCTTGTCTGGATCTTTTCTGCCACCACCAGCTCCTCCCGGA 497

Qy    476 gcaggaggtttgacggtgtatctagccggaggtcaaggacaagttgtaggagggaatgtg 535
          |||    | || ||| | ||||| || ||||| |||||||||||| || ||||| |  |||
Db    498 GCAACTAGCTTAACGATTTATCTTGCTGGAGGACAAGGACAAGTGGTGGGAGGTAGCGTG 557

Qy    536 gctggttcgttaattgcttcgggac 560
          | ||||  | | || ||  | || |
Db    558 GTTGGTGAGCTTATCGCGGCCGGGC 582
```

RESULT 15
BH566718/c
LOCUS BH566718 747 bp DNA linear GSS 14-DEC-2001
DEFINITION BOHCV23TR BOHC Brassica oleracea genomic clone BOHCV23, DNA sequence.
ACCESSION BH566718
VERSION BH566718.1 GI:17818558
KEYWORDS GSS.
SOURCE Brassica oleracea.
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 747)
AUTHORS Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished (2001)
COMMENT Other_GSSs: BOHCV23TF
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.
FEATURES Location/Qualifiers
source 1. .747
/organism="Brassica oleracea"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BOHCV23"
/clone_lib="BOHC"
/note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT 166 a 194 c 202 g 185 t
ORIGIN

Query Match 17.8%; Score 173.4; DB 12; Length 747;
Best Local Similarity 61.7%; Pred. No. 5.9e-37;
Matches 304; Conservative 0; Mismatches 171; Indels 18; Gaps 1;

```
Qy    156 ccggacgtcgtccacgtggtcgtcctgcaggttccaaaaacaaacccaaacctccgacga 215
          ||   || || |||||||| |||||||| |||||||| |||||||| || || || |
Db    508 CCACGCGCCGACCACGTGGACGTCCTGCTGGTTCCAAGAACAAACCAAAGCCACCAATCT 449

Qy    216 ttataactagagatagtcctaacgtccttagatcacacgttcttgaagtcacctccggtt 275
          || | ||  | |||  ||| ||||  || |     || ||  | ||  || |    ||
Db    448 TTGTCACGCGTGATTCTCCAAACGCTCTCAAGAGCCATGTCATGGAGATCGCTAGTGGGA 389

Qy    276 cggacatatccgaggcagtctccacctacgccactcgtcgcggctgcggcgtttgcatta 335
          | ||  |    ||  | |  | || | ||| |  || |||    | ||| | |||||
Db    388 CTGATGTCATAGAAACCCTAGCTACTTTCGCTAGGCGGCGCCAACGTGGCATCTGCATCT 329

Qy    336 taagcggcacgggtgcggtcactaacgtcacgatacggcaac---------- 377
          | ||||| |  ||  ||||  ||||||||||  | || ||||
Db    328 TGAGCGGTAACGGCACGGTGGCTAACGTCACACTCCGTCAACCATCAGTGGCTCCCGTTG 269

Qy    378 ctgcggctccggctggtggaggtgtgattaccctgcatggtcggtttgacattttgtctt 437
          | || || || | |||||  | |||  |  |  | || ||  ||||||| ||| | |||
Db    268 CAGCTGCCCCTGGTGGTGCGGCTGTATTGGCGTTACAAGGGAGGTTTGAGATTCTTTCTC 209

Qy    438 tgaccggtactgcgcttccaccgcctgcaccaccgggagcaggaggtttgacggtgtatc 497
          | ||||||  ||    | ||    || || ||||| ||| |    ||||| ||  | ||
Db    208 TAACCGGTTCTTTCTTACCTGGACCGGCTCCACCTGGATCCACTGGTTTAACTATTTACT 149

Qy    498 tagccggaggtcaaggacaagttgtaggagggaatgtggctggttcgttaattgcttcgg 557
          |||| || |||||||| || ||||| |||||  |  ||||  ||  | || || ||| | |
Db    148 TAGCTGGTGGTCAAGGTCAGGTTGTTGGAGGAAGCGTGGTGGGGGCATTGATGGCTGCTG 89

Qy    558 gaccggtagtgttgatggctgcttcttttgcaaacgcagtttatgataggttaccgattg 617
          | |||||  || | || |||||  | ||| | || ||   |||||| || |||||  | |
Db     88 GTCCGGTGATGCTAATCGCTGCCACGTTTTCTAATGCGACTTATGAGAGATTACCTTTGG 29

Qy    618 aagaggaagaaac 630
          | ||||||||| |
Db     28 ATGAGGAAGAAGC 16
```

Search completed: July 10, 2002, 20:39:48
Job time: 5921 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2002, 19:01:52 ; Search time 1922.93 Seconds
(without alignments)
10599.682 Million cell updates/sec

Title: US-09-823-676-1
Perfect score: 974
Sequence: 1 ccccccgacctgcctctaca.....cgaaatgcgaatattaggtt 974

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : [redacted]
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	952.8	97.8	86710	8	ATF23E12	AL022604 Arabidops
2	952.8	97.8	197859	8	ATCHRIV83	AL161587 Arabidops
3	212.4	21.8	1323	8	AF194974	AF194974 Arabidops
c 4	212.4	21.8	103192	8	AC007369	AC007369 Arabidops
5	208.6	21.4	88411	8	AC004667	AC004667 Arabidops
6	205.2	21.1	122871	8	AC002387	AC002387 Arabidops
7	201.8	20.7	100806	8	AC015450	AC015450 Arabidops
c 8	196.8	20.2	107600	8	ATF16J13	AL049638 Arabidops
c 9	196.8	20.2	190026	8	ATCHRIV33	AL161533 Arabidops
10	191.6	19.7	197976	8	ATCHRIV39	AL161539 Arabidops
11	191.6	19.7	206606	8	ATFCA1	Z97336 Arabidopsis
12	190.4	19.5	1161	8	AF446359	AF446359 Arabidops
c 13	190.4	19.5	81835	8	ATFCA9	Z97344 Arabidopsis
c 14	190.4	19.5	198067	8	ATCHRIV47	AL161547 Arabidops
15	186.4	19.1	89350	8	ATT12H17	AL021635 Arabidops
16	186.4	19.1	193550	8	ATCHRIV58	AL161558 Arabidops
c 17	183	18.8	129845	2	AP004165	AP004165 Oryza sat
18	180.2	18.5	177418	2	OSJN00182	AL662981 Oryza sat
19	179.4	18.4	107865	8	ATT4C21	AL162295 Arabidops
c 20	178.8	18.4	103240	8	ATT22E16	AL132975 Arabidops
21	177.2	18.2	95310	8	ATAC011437	AC011437 Arabidops
22	174	17.9	160363	2	AP003526	AP003526 Oryza sat
c 23	170.4	17.5	116205	8	AC006931	AC006931 Arabidops
c 24	166	17.0	156834	2	AP004757	AP004757 Oryza sat
25	161.8	16.6	134712	2	AP003891	AP003891 Oryza sat
26	161.8	16.6	135769	2	AP004587	AP004587 Oryza sat
27	160.8	16.5	108763	2	AP004020	AP004020 Oryza sat
c 28	148.8	15.3	147666	2	AP004635	AP004635 Oryza sat
29	148.8	15.3	152381	2	AP004698	AP004698 Oryza sat
c 30	127.6	13.1	12319	6	E64635	E64635 Method for
31	127.6	13.1	71807	8	AB025613	AB025613 Arabidops
32	116.8	12.0	145847	2	AP004319	AP004319 Oryza sat
33	111.6	11.5	140466	8	AP003683	AP003683 Oryza sat
c 34	108.2	11.1	111686	8	F14L17	AC012188 Sequence
35	101.2	10.4	279	8	AF184830	AF184830 Arabidops
c 36	89.2	9.2	140327	2	AP003938	AP003938 Oryza sat
c 37	87.8	9.0	169656	2	AP004680	AP004680 Oryza sat
38	69.8	7.2	167764	2	AP004303	AP004303 Oryza sat
c 39	50.8	5.2	125020	9	AF429315	AF429315 Homo sapi
40	45.2	4.6	590	8	HVU234403	AJ234403 Hordeum v
41	45.2	4.6	124050	8	AF474373	AF474373 Hordeum v
c 42	44.6	4.6	7218	6	I66494	I66494 Sequence 14
43	43.8	4.5	125020	9	AF429315	AF429315 Homo sapi
44	42.2	4.3	1137	8	AMA132349	AJ132349 Antirrhin
45	42	4.3	1694	8	ATAJ4119	AJ224119 Arabidops

ALIGNMENTS

RESULT 1
ATF23E12/c
LOCUS ATF23E12 86710 bp DNA linear PLN 01-APR-1999
DEFINITION Arabidopsis thaliana DNA chromosome 4, BAC clone F23E12 (ESSA project).
ACCESSION AL022604
VERSION AL022604.1 GI:3080406
KEYWORDS .
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 86710)
AUTHORS Bevan,M., Hilbert,H., Braun,M., Holzer,E., Brandt,A., Duesterhoeft,A., Hoheisel,J., Mewes,H.W., Mayer,K.F.X. and Schueller,C.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 86710)
AUTHORS EU Arabidopsis sequencing,project.
TITLE Direct Submission
JOURNAL Submitted (01-APR-1999) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:

```
            schuelle@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
COMMENT     Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
            and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.
FEATURES              Location/Qualifiers
     source           1. .86710
                      /organism="Arabidopsis thaliana"
                      /variety="Columbia"
                      /db_xref="taxon:3702"
                      /chromosome="4"
     gene             173. .1890
                      /gene="F23E12.10"
     exon             173. .374
                      /gene="F23E12.10"
                      /number=1
     CDS              join(<173. .374,456. .919,1089. .1610,1678. .1890)
                      /gene="F23E12.10"
                      /codon_start=1
                      /product="hypothetical protein (fragment)"
                      /protein_id="CAA18726.1"
                      /db_xref="GI:3080407"
                      /db_xref="SPTREMBL:O65486"
                      /translation="FNAAVAGCFFAVESVLWPSSSTDSSTSLPNTTSMVILSAVTASV
                      VSEIGLGSEPAFKVPDYDFRSPGELPLYLLLGALCGLVSLALSRCTSSMTSAVDSLNK
                      DAGIPKAVFPVMGGLSVGIIALVYPEVLYWGFQNVDILLEKRPFVKGLSADLLLQLVA
                      VKIAATAWCRASGLVGGYYAPSLFIGGAAGMAYGKFIGLALAQNPDFNLSILEVASPQ
                      AYGLVGMAATLAGVCQVPLTAVLLLFELTQDYRIVLPLLGAVGMSSWITSGQSKRQET
                      RETKETRKRKSQEAVQSLTSSDDESSTNNLCEVESSLCLDDSLNQSEELPKSIFVSEA
                      MRTRFATVMMSTSLEEALTRMLIEKQSCALIVDPDNIFLGILTLSDIQEFSKARKEGN
                      NRPKDIFVNDICSRSGGKCKVPWTVTPDMDLLAAQTIMNKHELSHVAVVSGSIDAPRI
                      HPVGVLDRECITLTRR"
     intron           375. .455
                      /gene="F23E12.10"
                      /number=1
     exon             456. .919
                      /gene="F23E12.10"
                      /number=2
     intron           920. .1088
                      /gene="F23E12.10"
                      /number=2
     exon             1089. .1610
                      /gene="F23E12.10"
                      /number=3
     intron           1611. .1677
                      /gene="F23E12.10"
                      /number=3
     exon             1678. .1890
                      /gene="F23E12.10"
                      /number=4
     exon             3124. .3241
                      /gene="F23E12.20"
                      /number=1
     gene             3124. .4428
                      /gene="F23E12.20"
     CDS              join(3124. .3241,3323. .3465,3595. .3696,3945. .4107,
                      4211. .4428)
                      /gene="F23E12.20"
                      /note="similarity to dihydroflavonol-4-reductase, Gerbera
                      x sp., PIR2:S35189"
                      /codon_start=1
                      /product="putative protein"
                      /protein_id="CAA18727.1"
                      /db_xref="GI:3080408"
                      /db_xref="SPTREMBL:O65487"
                      /translation="MDQAKGKVCVTGASGFLASWLVKRLLLEGYEVIGTVRDPGNEKK
                      LAHLWKLEGAKERLRLVKADLMEEGSFDNAIMGCQGVFHTASPEILRPAIEGTLNVLR
                      SCRKNPSLKRVVLTSSSSTVWYALSKTLAEQAAWKFSEENGIDLVTVLPSFLVGPSLP
                      PDLCSTASDVLGLLKGETEKFQWHGQMGYVHIDDVARTHIVVFEHEAAQGRYICSSNV
                      ISLEELVSFLSARYPSLPIPKRFDTDPLN"
     intron           3242. .3322
                      /gene="F23E12.20"
                      /number=1
     exon             3323. .3465
                      /gene="F23E12.20"
                      /number=2
     intron           3466. .3594
                      /gene="F23E12.20"
                      /number=2
     exon             3595. .3696
                      /gene="F23E12.20"
                      /number=3
     intron           3697. .3944
                      /gene="F23E12.20"
                      /number=3
     exon             3945. .4107
                      /gene="F23E12.20"
                      /number=4
     intron           4108. .4210
                      /gene="F23E12.20"
                      /number=4
     exon             4211. .4428
                      /gene="F23E12.20"
                      /number=5
     gene             complement(join(4939. .5013,5109. .5195,5307. .5342,
                      5451. .5651,5725. .5811,6161. .6163))
                      /gene="F23E12.30"
     gene             4939. .6163
                      /gene="F23E12.30"
     CDS              complement(join(4939. .5013,5109. .5195,5307. .5342,
                      5451. .5651,5725. .5811,6161. .6163))
                      /gene="F23E12.30"
                      /note="Contains Clathrin adaptor complexes small chain
                      signature [VVYKRYASLYF]
                      contains EST gb:T44219"
                      /codon_start=1
                      /product="clathrin assembly protein AP19 homolog"
                      /protein_id="CAA18728.1"
                      /db_xref="GI:3080409"
                      /db_xref="SPTREMBL:O23685"
                      /translation="MIHFVLLVSRQGKVRLTKWYSPYAQKERSKVIRELSGVILNRGP
                      KLCNFVEWRGYKVVYKRYASLYFCMCIDQEDNELEVLEIIHHYVEILDRYFGSVCELD
                      LIFNFHKAYYILDELLIAGELQESSKKTVARIISAQDQLVEVAKEEASSISNIIAQAT
                      NR"
     exon             complement(4939. .5013)
                      /gene="F23E12.30"
                      /number=1
     intron           complement(5014. .5108)
                      /gene="F23E12.30"
                      /number=1
     exon             complement(5109. .5195)
                      /gene="F23E12.30"
                      /number=2
     intron           complement(5196. .5306)
                      /gene="F23E12.30"
                      /number=2
     exon             complement(5307. .5342)
                      /gene="F23E12.30"
                      /number=3
     intron           complement(5343. .5450)
                      /gene="F23E12.30"
                      /number=3
     exon             complement(5451. .5651)
                      /gene="F23E12.30"
                      /number=4
     intron           complement(5652. .5724)
                      /gene="F23E12.30"
                      /number=4
     exon             complement(5725. .5811)
                      /gene="F23E12.30"
                      /number=5
     intron           complement(5812. .6160)
                      /gene="F23E12.30"
                      /number=5
```

```
     exon            complement(6161. .6163)
                     /gene="F23E12.30"
     exon            /number=6
                     complement(7352. .7582)
                     /gene="F23E12.40"
     gene            /number=1
                     7352. .7582
     gene            /gene="F23E12.40"
                     complement(7352. .7582)
     CDS             /gene="F23E12.40"
                     complement(7352. .7582)
                     /gene="F23E12.40"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAA18729.1"
                     /db_xref="GI:3080410"
                     /db_xref="SPTREMBL:O65488"
                     /translation="MPLVRKNGVVMRIGSPMLHESENLVGDGVEAPPRGERDEIGEEV
     gene            GGETNVVSESCGGEEGAQDAAKTTPEGEADRI"
                     8233. .9045
     gene            /gene="F23E12.50"
                     complement(8233. .9045)
     exon            /gene="F23E12.50"
                     complement(8233. .9045)
                     /gene="F23E12.50"
     CDS             /number=1
                     complement(8233. .9045)
                     /gene="F23E12.50"
                     /note="similarity to DNA-binding protein PD1, Pisum
                     sativum"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAA18730.1"

  Query Match             97.8%;  Score 952.8;  DB 8;  Length 86710;
  Best Local Similarity   99.7%;  Pred. No. 2.1e-258;
  Matches  965;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy      7 gacctgcctctacagagacctgaagattccagaaccccacctgatcaaaaataacatgga 66
          || |||| ||||||||||||||||||||||||||||||||||| ||||||||||||
Db   9099 GAACTGCATCTACAGAGACCTGAAGATTCCAGAACCCCACCTGATC-AAAATAACATGGA 9041

Qy     67 acttaacagatctgaagcagacgaagcaaaggccgagaccactcccaccggtggagccac 126
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9040 ACTTAACAGATCTGAAGCAGACGAAGCAAAGGCCGAGACCACTCCCACCGGTGGAGCCAC 126

Qy    127 cagctcagccacagcctctggctcttcctccggacgtcgtccacgtggtcgtcctgcagg 186
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8980 CAGCTCAGCCACAGCCTCTGGCTCTTCCTCCGGACGTCGTCCACGTGGTCGTCCTGCAGG 8921

Qy    187 ttccaaaaacaaacccaaacctccgacgattataactagagatagtcctaacgtccttag 246
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8920 TTCCAAAAACAAACCCAAACCTCCGACGATTATAACTAGAGATAGTCCTAACGTCCTTAG 8861

Qy    247 atcacacgttcttgaagtcacctccggttcggacatatccgaggcagtctccacctacgc 306
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8860 ATCACACGTTCTTGAAGTCACCTCCGGTTCGGACATATCCGAGGCAGTCTCCACCTACGC 8801

Qy    307 cactcgtcgcggctgcggcgtttgcattataagcggcacgggtgcggtcactaacgtcac 366
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8800 CACTCGTCGCGGCTGCGGCGTTTGCATTATAAGCGGCACGGGTGCGGTCACTAACGTCAC 8741

Qy    367 gatacggcaacctgcggctccggctggtggaggtgtgattaccctgcatggtcggtttga 426
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8740 GATACGGCAACCTGCGGCTCCGGCTGGTGGAGGTGTGATTACCCTGCATGGTCGGTTTGA 8681

Qy    427 cattttgtctttgaccggtactgcgcttccaccgcctgcaccaccgggagcaggaggttt 486
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8680 CATTTTGTCTTTGACCGGTACTGCGCTTCCACCGCCTGCACCACCGGGAGCAGGAGGTTT 8621

Qy    487 gacggtgtatctagccggaggtcaaggacaagttgtaggagggaatgtggctggttcgtt 546
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8620 GACGGTGTATCTAGCCGGAGGTCAAGGACAAGTTGTAGGAGGGAATGTGGCTGGTTCGTT 8561

Qy    547 aattgcttcgggaccggtagtgttgatggctgcttcttttgcaaacgcagtttatgatag 606
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8560 AATTGCTTCGGGACCGGTAGTGTTGATGGCTGCTTCTTTTGCAAACGCAGTTTATGATAG 8501

Qy    607 gttaccgattgaagaggaagaaaccccaccgccgagaaccaccgggggtgcagcagcagca 666
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8500 GTTACCGATTGAAGAGGAAGAAACCCCACCGCCGAGAACCACCGGGGTGCAGCAGCAGCA 8441

Qy    667 gccggaggcgtctcagtcgtcggaggttacggggagtggggcccaggcgtgtgagtcaaa 726
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8440 GCCGGAGGCGTCTCAGTCGTCGGAGGTTACGGGGAGTGGGGCCCAGGCGTGTGAGTCAAA 8381

Qy    727 cctccaaggtggaaatggtggaggaggtgttgctttctacaatcttggaatgaatatgaa 786
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8380 CCTCCAAGGTGGAAATGGTGGAGGAGGTGTTGCTTTCTACAATCTTGGAATGAATATGAA 8321

Qy    787 caattttcaattctccgggggagatatttacggtatgagcggcggtagcggaggaggtgg 846
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8320 CAATTTTCAATTCTCCGGGGGAGATATTTACGGTATGAGCGGCGGTAGCGGAGGAGGTGG 8261

Qy    847 tggcggtgcgactagacccgcgttttagagttttagcgttttggtgacacctttgttgc 906
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8260 TGGCGGTGCGACTAGACCCGCGTTTTAGAGTTTTAGCGTTTTGGTGACACCTTTTGTTGC 8201

Qy    907 gtttgcgtgtttgacctcaaactactaggctactagctatagcggttgcgaaatgcgaat 966
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   8200 GTTTGCGTGTTTGACCTCAAACTACTAGGCTACTAGCTATAGCGGTTGCGAAATGCGAAT 8141

Qy    967 attaggtt 974
          ||||||||
Db   8140 ATTAGGTT 8133


RESULT   2
ATCHRIV83
LOCUS       ATCHRIV83               197859 bp    DNA     linear   PLN 16-MAR-2000
DEFINITION  Arabidopsis thaliana DNA chromosome 4, contig fragment No. 83.
ACCESSION   AL161587
VERSION     AL161587.2  GI:7270470
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 16641 to 103350)
  AUTHORS   Hilbert,H., Braun,M., Holzer,E., Brandt,A., Duesterhoeft,A.,
            Mewes,H.W., Lemcke,K. and Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   2  (bases 98352 to 125759)
  AUTHORS   Robben,J., Grymonprez,B., Volckaert,G., Mewes,H.W., Lemcke,K. and
            Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   3  (bases 120761 to 197859)
  AUTHORS   Rose,M., Hempel,S., Entian,K.-D., Mewes,H.W., Lemcke,K. and
            Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   4  (bases 1 to 24256)
  AUTHORS   Murphy,G., Ridley,P., Hudson,S., Mewes,H.W., Lemcke,K. and
            Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   5  (bases 1 to 197859)
  AUTHORS   EU Arabidopsis sequencing,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
COMMENT     Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2002, 22:19:08 ; Search time 103.68 Seconds
(without alignments)
450.508 Million cell updates/sec

Title: US-09-823-676-2
Perfect score: 1392
Sequence: 1 MELNRSEADEAKAETTPTGG.....IYGMSGGSGGGGGGATRPAF 270

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1392	100.0	270	10	O65489	O65489 arabidopsis
2	706.5	50.8	311	10	Q9S7C9	Q9s7c9 arabidopsis
3	646.5	46.4	292	10	O23620	O23620 arabidopsis
4	645.5	46.4	302	10	Q9C9K7	Q9c9k7 arabidopsis
5	637.5	45.8	315	10	Q9SR17	Q9sr17 arabidopsis
6	629.5	45.2	317	10	O22130	O22130 arabidopsis
7	625.5	44.9	339	10	Q9SZ70	Q9sz70 arabidopsis
8	616	44.3	285	10	O82166	O82166 arabidopsis
9	605	43.5	324	10	O49662	O49662 arabidopsis
10	581	41.7	265	10	Q9LZX7	Q9lzx7 arabidopsis
11	572	41.1	310	10	Q9M2S3	Q9m2s3 arabidopsis
12	554	39.8	257	10	Q9SJG4	Q9sjg4 arabidopsis
13	499.5	35.9	276	10	Q9LTA2	Q9lta2 arabidopsis
14	417	30.0	206	10	Q9M9R4	Q9m9r4 arabidopsis
15	407	29.2	265	10	Q941Y9	Q941y9 oryza sativ
16	289.5	20.8	404	10	Q9LVB0	Q9lvb0 arabidopsis
17	289	20.8	771	10	Q9XHY4	Q9xhy4 oryza sativ
18	283	20.3	405	10	Q9AYM0	Q9aym0 oryza sativ
19	282	20.3	351	10	Q94F52	Q94f52 arabidopsis
20	274	19.7	296	10	O22812	O22812 arabidopsis
21	274	19.7	351	10	O23142	O23142 arabidopsis
22	264.5	19.0	419	10	Q9FHM5	Q9fhm5 arabidopsis
23	259	18.6	826	10	Q9SH33	Q9sh33 arabidopsis
24	258.5	18.6	300	10	Q9ZRR7	Q9zrr7 antirrhinum
25	255.5	18.4	334	10	O49658	O49658 arabidopsis
26	248.5	17.9	386	10	Q9FIR1	Q9fir1 arabidopsis
27	246.5	17.7	347	10	O04696	O04696 pisum sativ
28	243	17.5	439	10	O49350	O49350 arabidopsis
29	243	17.5	439	10	Q940I0	Q940i0 arabidopsis
30	242.5	17.4	334	10	O04695	O04695 pisum sativ
31	239.5	17.2	404	10	Q9SB31	Q9sb31 arabidopsis
32	239	17.2	455	10	O49694	O49694 arabidopsis
33	227	16.3	365	10	Q9SZ73	Q9sz73 arabidopsis
34	225	16.2	348	10	O80834	O80834 arabidopsis
35	221	15.9	418	10	Q9SR16	Q9sr16 arabidopsis
36	218	15.7	345	10	O81321	O81321 arabidopsis
37	211.5	15.2	348	10	Q9M2D3	Q9m2d3 arabidopsis
38	162.5	11.7	574	10	Q9SJQ2	Q9sjq2 arabidopsis
39	148	10.6	848	5	Q9V889	Q9v889 drosophila
40	146	10.5	405	5	Q23057	Q23057 caenorhabdi
41	143.5	10.3	818	6	Q9N1P0	Q9n1p0 bos taurus
42	139	10.0	1468	5	Q9GUB5	Q9gub5 galleria me
43	138.5	9.9	265	5	Q23347	Q23347 caenorhabdi
44	137	9.8	694	16	O53212	O53212 mycobacteri
45	136.5	9.8	1610	16	Q92KQ8	Q92kq8 rhizobium m

ALIGNMENTS

RESULT 1
O65489
ID O65489 PRELIMINARY; PRT; 270 AA.
AC O65489;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 26.8 KDA PROTEIN.
GN F23E12.50 OR AT4G35390.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA Hoheisel J., Jesse T., Heijnen L., Vos P., Mewes H.W., Mayer K.F.X.,
RA Schueller C.;
RL Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AL022604; CAA18730.1; -.
DR EMBL; AL161587; CAB80256.1; -.
KW Hypothetical protein.
SQ SEQUENCE 270 AA; 26774 MW; 4C21487C2719E114 CRC64;

Query Match 100.0%; Score 1392; DB 10; Length 270;
Best Local Similarity 100.0%; Pred. No. 4.7e-88;
Matches 270; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MELNRSEADEAKAETTPTGGATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDSPNV 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MELNRSEADEAKAETTPTGGATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDSPNV 60

Qy      61 LRSHVLEVTSGSDISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGR 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 LRSHVLEVTSGSDISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGR 120

Qy     121 FDILSLTGTALPPPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVY 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 FDILSLTGTALPPPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVY 180

Qy     181 DRLPIEEEETPPPRTTGVQQQQPEASQSSEVTGSGAQACESNLQGGNGGGGVAFYNLGMN 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 DRLPIEEEETPPPRTTGVQQQQPEASQSSEVTGSGAQACESNLQGGNGGGGVAFYNLGMN 240

Qy     241 MNNFQFSGGDIYGMSGGSGGGGGGATRPAF 270
           ||||||||||||||||||||||||||||||
Db     241 MNNFQFSGGDIYGMSGGSGGGGGGATRPAF 270
```

```
ULT   2
7C9
ID   Q9S7C9      PRELIMINARY;      PRT;   311 AA.
AC   Q9S7C9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   ESCAROLA.
GN   ESC OR F9H16.12.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Weigel D., Ahn J.H., Blazquez M.A., Borevitz J., Christensen S.K.,
RA   Fankhauser C., Ferrandiz C., Kardailsky I., Neff M.M., Nguyen J.T.,
RA   Sato S., Wang Z., Xia Y., Dixon R.A., Harrison M.J., Lamb C.J.,
RA   Yanofsky M.F., Chory J.;
RT   "Activation Tagging in Arabidopsis.";
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,
RA   Altafi H., Araujo R., Huizar L., Rowley D., Buehler E., Dunn P.,
RA   Gonzalez A., Kremenetskaia I., Kim C., Lenz C., Li J., Liu S.,
RA   Luros S., Schwartz J., Shinn P., Toriumi M., Vysotskaia V.S.,
RA   Walker M., Yu G., Ecker J., Theologis A., Davis R.W.;
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF194974; AAF07197.1; -.
DR   EMBL; AC007369; AAD30602.1; -.
DR   InterPro; IPR000637; AT_hook.
DR   Pfam; PF02178; AT_hook; 1.
DR   SMART; SM00384; AT_hook; 1.
SQ   SEQUENCE   311 AA;  31842 MW;  A80B445C9776EB7D CRC64;
```

```
  Query Match           50.8%;  Score 706.5;  DB 10;  Length 311;
  Best Local Similarity   53.0%;  Pred. No. 5e-41;
  Matches  151;  Conservative   25;  Mismatches   56;  Indels   53;  Gaps    6;

Qy       7 EADEAKAETTPTGGATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDSPNVLRSHVL 66
           |:| :  :    |   |    | |: |:|||||| ||||| ||| |:|||||| ||||||
Db      59 ESDHSNKDHHQQGRPDSDPNTSSSAPGKRPRGRPPGSKNKAKPPIIVTRDSPNALRSHVL 118

Qy      67 EVTSGSDISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAP-------AGGGVITLHG 119
           ||: |:|| |:||||| ||| || :: | | |:|||:|||  |        ||||:|||||
Db     119 EVSPGADIVESVSTYARRRGRGVSVLGGNGTVSNVTLRQPVTPGNGGGVSGGGGVVTLHG 178

Qy     120 RFDILSLTGTALPPPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAV 179
```

```
           ||:|||||| |||||||||||:::|||||||||||:|   |||| ||:||||||:|||
Db     179 RFEILSLTGTVLPPPAPPGAGGLSIFLAGGQGQVVGGSVVAPLIASAPVILMAASFSNAV 238

Qy     180 YDRLPIEEEEETP-----------PPRTTGVQQQQPEASQSSEVTGSGAQACESNLQGGNG 228
           ::||||||||              ||:    || | ||  | ||| |          |||
Db     239 FERLPIEEEEEEGGGGGGGGGGGPPQ----MQQAPSASPPSGVTGQGQ-------LGGNV 287

Qy     229 GGGVAFYNLGMNMNNFQFSGGDIYGMSGGS---GGGGGGATRPAF 270
           ||                    || ||     | | |  :|| |
Db     288 GG--------------------YGFSGDPHLLGWGAGTPSRPPF 311
```

```
RESULT   3
O23620
ID   O23620      PRELIMINARY;      PRT;   292 AA.
AC   O23620;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 29.7 KDA PROTEIN.
GN   AT4G17800.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98121113; PubMed=9461215;
RA   Bevan M., Bancroft I., Bent E., Love K., Goodman H., Dean C.,
RA   Bergkamp R., Dirkse W., van Staveren M., Stiekema W., Drost L.,
RA   Ridley P., Hudson S.A., Patel K., Murphy G., Piffanelli P., Wedler H.,
RA   Wedler E., Wambutt R., Weitzenegger T., Pohl T.M., Terryn N.,
RA   Gielen J., Villarroel R., De Clerck R., van Montagu M., Lecharny A.,
RA   Auborg S., Gy I., Kreis M., Lao N., Kavanagh T., Hempel S., Kotter P.,
RA   Entian K.D., Rieger M., Schaeffer M., Funk B., Mueller-Auer S.,
RA   Silvey M., James R., Montfort A., Pons A., Puigdomenech P., Douka A.,
RA   Voukelatou E., Milioni D., Hatzopoulos P., Piravandi E., Obermaier B.,
RA   Hilbert H., Duesterhoft A., Moores T., Jones J.D.G., Eneva T.,
RA   Palme K., Benes V., Rechman S., Ansorge W., Cooke R., Berger C.,
RA   Delseny M., Voet M., Volckaert G., Mewes H.W., Klosterman S.,
RA   Schueller C., Chalwatzis N.;
RT   "Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of
RT   Arabidopsis thaliana.";
RL   Nature 391:485-488(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z97344; CAB10560.1; -.
DR   EMBL; AL161547; CAB78783.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   292 AA;  29724 MW;  2A948700FA543C4F CRC64;
```

```
  Query Match           46.4%;  Score 646.5;  DB 10;  Length 292;
  Best Local Similarity   52.7%;  Pred. No. 5.9e-37;
  Matches  136;  Conservative   32;  Mismatches   47;  Indels   43;  Gaps    8;

Qy      19 GGATSSATASGSSS-----GRRPRGRPAGSKNKPKPPTIITRDSPNVLRSHVLEVTSGSD 73
           ||: ||    |        |||||||| |||||||| ||||:| | ||:|:||||:| |
Db      62 GGSGSSGGGGGHGGGGDVVGRRPRGRPPGSKNKPKPPVIITRESANTLRAHILEVTNGCD 121

Qy      74 ISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGRFDILSLTGTALPP 133
           : : |:||| ||  |:|::||:| ||||:||||:  | | |:|| | |:|||:|: |||
Db     122 VFDCVATYARRRQRGICVLSGSGTVTNVSIRQPS--AAGAVVTLQGTFEILSLSGSFLPP 179

Qy     134 PAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVYDRLPIEEEETPPP 193
           ||||||  ||::||||||||||:| | | |:|||:::|||| |  |:|||:||:|
Db     180 PAPPGATSLTIFLAGGQGQVVGGSVVGELTAAGPVIVIAASFTNVAYERLPLEEDE---- 235

Qy     194 RTTGVQQQQPEASQSSEVTGSGAQACESNLQGGN--------GGGGVAFYNLGMNMN-NF 244
               ||||          | |:        |||        ||||: |:|| |||  |
```

PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.

PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 45.8%; Score 637.5; DB 21; Length 280;
Best Local Similarity 50.0%; Pred. No. 9.4e-44;
Matches 132; Conservative 37; Mismatches 58; Indels 37; Gaps 5;

```
Qy     17 PTGGATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDSPNVLRSHVLEVTSGSDISE 76
          |  ||  : |       ||||||||||||||||  :|||||| |:|||:|: ||:|: |
Db     35 pregaveapt-------rrprgrpagsknkpkppifvtrdspnalkshvmeiasgtdvie 87

Qy     77 AVSTYATRRGCGVCIISGTGAVTNVTIRQP------AAPAGGGVITLHGRFDILSLTGTA 130
           ::|:| ||  |:|||:|| | | |||:|||      ||| |  |: | |||:||||||:
Db     88 tlatfarrrqrgicilsgngtvanvtlrqpstaavaaapggaavlalqgrfeilsltgsf 147

Qy    131 LPPPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVYDRLPIEEEET 190
          || |||||: |||:|||||||||||||:| | |:|:|||:|:||:|:|| |:|||:||||
Db    148 lpgpappgstgltiylaggqgqvvggsvvgplmaagpvmliaatfsnatyerlpleeeea 207

Qy    191 PP----PRTTGVQQQQPEASQSSEVTGSGAQACESNLQGGNGGGGVAFYNLGMNMNNFQF 246
                : ||   |          | |     |   ||:|  |:  ||:  |: :
Db    208 aergggggsggvvpgq---------lgggsplssgagggdgnqglpvynmpgnlvs--- 255

Qy    247 SGGDIYGMSGGSGGGGGGATRPAF 270
                :|||||||  : : |:
Db    256 --------nggsgggqmsgqeay 271
```

RESULT 6
AAG10649
ID AAG10649 standard; Protein; 282 AA.
XX
AC AAG10649;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 9055.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.

```
Qy      19 GGATSSATASGSSS-----GRRPRGRPAGSKNKPKPPTIITRDSPNVLRSHVLEVTSGSD 73
           ||: ||    |        |||||||| |||||||| ||||:| | ||:|:||||:| |
Db      62 ggsgssgggggghggggdvvgrrprgrppgsknkpkppviitresantlrahilevtngcd 121

Qy      74 ISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGRFDILSLTGTALPP 133
           : : |:||| ||  |:|::||:| ||||:||||:  | | |:|| | |:||||:|: |||
Db     122 vfdcvatyarrrqrgicvlsgsgtvtnvsirqps--aagavvtlqgtfeilslsgsflpp 179

Qy     134 PAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVYDRLPIEEEETPPP 193
           ||||||  ||::||||||||||||:| | | |:|||:::|||| |  |:|||:||:|
Db     180 pappgatsltiflaggqgqvvggsvvgeltaagpviviaasftnvayerlpleede---- 235

Qy     194 RTTGVQQQQPEASQSSEVTGSGAQACESNLQGGN--------GGGGVAFYNLGMNMN-NF 244
                ||||          | |:        |||        ||||: |:|| |||  |
Db     236 -----qqqq---------lgggsng------ggnlfpevaaggggglpffnlpmnmqpnv 275

Qy     245 QFSGGDIYGMSGGSGGGG 262
           |     : |  | ||| |
Db     276 ql---pvegwpgnsggrg 290


RESULT   5
AAG10650
ID   AAG10650 standard; Protein; 280 AA.
XX
AC   AAG10650;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 9056.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   20-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0144814.
PR   21-JUL-1999;   99US-0145086.
PR   21-JUL-1999;   99US-0145088.
PR   22-JUL-1999;   99US-0145085.
PR   22-JUL-1999;   99US-0145087.
PR   22-JUL-1999;   99US-0145089.
PR   22-JUL-1999;   99US-0145192.
PR   23-JUL-1999;   99US-0145145.
PR   23-JUL-1999;   99US-0145218.
PR   23-JUL-1999;   99US-0145224.
PR   26-JUL-1999;   99US-0145276.
PR   27-JUL-1999;   99US-0145913.
PR   27-JUL-1999;   99US-0145918.
PR   27-JUL-1999;   99US-0145919.
PR   28-JUL-1999;   99US-0145951.
PR   02-AUG-1999;   99US-0146386.
```

PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 46.4%; Score 646.5; DB 21; Length 292;
Best Local Similarity 52.7%; Pred. No. 1.9e-44;
Matches 136; Conservative 32; Mismatches 47; Indels 43; Gaps 8;

```
PR   18-AUG-1999;   99US-0149426.
PR   20-AUG-1999;   99US-0149722.
PR   20-AUG-1999;   99US-0149723.
PR   20-AUG-1999;   99US-0149929.
PR   23-AUG-1999;   99US-0149902.
PR   23-AUG-1999;   99US-0149930.
PR   25-AUG-1999;   99US-0150566.
PR   26-AUG-1999;   99US-0150884.
PR   27-AUG-1999;   99US-0151065.
PR   27-AUG-1999;   99US-0151066.
PR   27-AUG-1999;   99US-0151080.
PR   30-AUG-1999;   99US-0151303.
PR   31-AUG-1999;   99US-0151438.
PR   01-SEP-1999;   99US-0151930.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match           46.4%;  Score 646.5;  DB 21;  Length 254;
  Best Local Similarity  52.7%;  Pred. No. 1.6e-44;
  Matches  136;  Conservative   32;  Mismatches   47;  Indels   43;  Gaps    8;

Qy      19 GGATSSATASGSSS-----GRRPRGRPAGSKNKPKPPTIITRDSPNVLRSHVLEVTSGSD 73
           ||: ||    |        |||||||| |||||||| ||||:| | ||:|:|||||:| |
Db      24 ggsgssgggggghggggdvvgrrprgrppgsknkpkppviitresantlrahilevtngcd 83

Qy      74 ISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGRFDILSLTGTALPP 133
           : : |:||| ||  |:|::||:| ||||:|||:  | | |:|| | |:|||:|: |||
Db      84 vfdcvatyarrrqrgicvlsgsgtvtnvsirqps--aagavvtlqgtfeilslsgsflpp 141

Qy     134 PAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVYDRLPIEEEETPPP 193
           ||||||  ||::|||||||||||:| | | |:|||:::|||| |  |:|||:||:|
Db     142 pappgatsltiflaggqgqvvggsvvgeltaagpviviaasftnvayerlpleede---- 197

Qy     194 RTTGVQQQQPEASQSSEVTGSGAQACESNLQGGN--------GGGGVAFYNLGMNMN-NF 244
           ||||          | |:        |||         ||||: |:|| |||  |
Db     198 -----qqqq---------lgggsng------ggnlfpevaaggggglpffnlpmnmqpnv 237

Qy     245 QFSGGDIYGMSGGSGGGG 262
           |     : |  | ||| |
Db     238 ql---pvegwpgnsggrg 252


RESULT   4
AAG29344
ID   AAG29344 standard; Protein; 292 AA.
XX
AC   AAG29344;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 34899.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
```

Db 241 mnnfqfsggdiygmsggsgggggggatrpaf 270

RESULT 3
AAG29345
ID AAG29345 standard; Protein; 254 AA.
XX
AC AAG29345;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 34900.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.

```
PA   (KEDD/) KEDDIE J.
PA   (RATC/) RATCLIFFE O.
PA   (HEAR/) HEARD J.
PA   (SAMA/) SAMAHA R.
PA   (YUGG/) YU G.
PA   (JIAN/) JIANG C.
XX
PI   Riechmann JL,  Reuber L,  Keddie J,  Ratcliffe O,  Heard J,  Samaha R;
PI   Yu G,  Jiang C;
XX
DR   WPI; 2001-335996/35.
DR   N-PSDB; AAD05839.
XX
PT   Nucleic acids encoding plant transcription factor polypeptides, useful
PT   for altering the developmental and structural characteristics of
PT   plants, e.g. corn, potato and cotton plants -
XX
PS   Claim 4; Page 103-104; 140pp; English.
XX
CC   The present sequence is Arabidopsis thaliana transcription factor G1073.
CC   The transcription factors are used to modify traits associated
CC   with structural or developmental characteristics of plants, e.g.soybean,
CC   wheat, corn, potato, cotton, rice, oilseed rape, sunflower, alfalfa,
CC   sugar cane, turf, banana, blackberry, blueberry, strawberry, raspberry,
CC   cantaloupe, carrot, cauliflower, coffee, cucumber, eggplant, grapes,
CC   honey dew, lettuce, mango, melon, papaya, peas, watermelon, pineapple,
CC   spinach, squash, sweet corn, tobacco, tomato, peppers, onion, rosaceous
CC   fruits and/or vegetable brassicas when their expression levels are
CC   altered. The manipulation of transcription factor levels in plants
CC   offers great potential in agricultural biotechnology for modifying
CC   plant's traits. The transcription factors are also used in gene therapy.
XX
SQ   Sequence   270 AA;


  Query Match             100.0%;  Score 1392;  DB 22;  Length 270;
  Best Local Similarity   100.0%;  Pred. No. 8.6e-105;
  Matches  270;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MELNRSEADEAKAETTPTGGATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDSPNV 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 melnrseadeakaettptggatssatasgsssgrrprgrpagsknkpkpptiitrdspnv 60

Qy      61 LRSHVLEVTSGSDISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGR 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 lrshvlevtsgsdiseavstyatrrgcgvciisgtgavtnvtirqpaapagggvitlhgr 120

Qy     121 FDILSLTGTALPPPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVY 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 fdilsltgtalpppappgaggltvylaggqgqvvggnvagsliasgpvvlmaasfanavy 180

Qy     181 DRLPIEEEETPPPRTTGVQQQQPEASQSSEVTGSGAQACESNLQGGNGGGGVAFYNLGMN 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 drlpieeeetppprttgvqqqqpeasqssevtgsgaqacesnlqggngggggvafynlgmn 240

Qy     241 MNNFQFSGGDIYGMSGGSGGGGGGATRPAF 270
           |||||||||||||||||||||||||||||
Db     241 mnnfqfsggdiygmsggsgggggggatrpaf 270


RESULT   2
AAB67838
ID   AAB67838 standard; Protein; 270 AA.
XX
AC   AAB67838;
XX
DT   29-JUN-2001  (first entry)
XX
DE   Amino acid sequence of a plant transcription factor G1073.
XX
KW   Transcription factor; flowering time; transgenic plant; vernalisation;
KW   plant development; plant physiology; flowering.
XX
OS   Arabidopsis thaliana.
XX
PN   WO200126459-A2.
XX
PD   19-APR-2001.
XX
PF   11-OCT-2000; 2000WO-US28141.
XX
PR   12-OCT-1999;   99US-0159464.
PR   08-NOV-1999;   99US-0164132.
PR   17-NOV-1999;   99US-0166228.
PR   17-APR-2000;  2000US-0197899.
PR   22-AUG-2000;  2000US-0227439.
XX
PA   (MEND-) MENDEL BIOTECHNOLOGY INC.
PA   (RATC/) RATCLIFFE O.
PA   (HEAR/) HEARD J.
PA   (SAMA/) SAMAHA R.
PA   (CREE/) CREELMAN R.
PA   (KEDD/) KEDDIE J.
PA   (JIAN/) JIANG C.
PA   (REUB/) REUBER L.
PA   (RIEC/) RIECHMANN J L.
XX
PI   Ratcliffe O,  Heard J,  Samaha R,  Creelman R,  Keddie J,  Jiang C;
PI   Reuber L,  Riechmann JL;
XX
DR   WPI; 2001-266398/27.
DR   N-PSDB; AAF80408.
XX
PT   New transgenic plant comprises a recombinant polynucleotide encoding a
PT   plant transcription factor polypeptide and has a modified flowering
PT   time or vernalization requirement -
XX
PS   Example 7; Page 94-95; 108pp; English.
XX
CC   The present sequence represents a plant transcription factor protein
CC   which modifies the flowering time of a plant. The polynucleotide
CC   sequence is used to produce transgenic plants which have a modified
CC   flowering time or a modified vernalisation requirement. The
CC   polynucleotides and polypeptides are useful for modifying plant
CC   development, physiology or biochemistry such that the modified plants
CC   have a trait advantage over wild type plants. In particular they are
CC   useful for accelerating, delaying or preventing flowering. The
CC   polynucleotides are also useful as nucleic acid probes and primers.
CC   They may be used to identify proteins that can modify the activity of
CC   the transcription factor.
XX
SQ   Sequence   270 AA;


  Query Match             100.0%;  Score 1392;  DB 22;  Length 270;
  Best Local Similarity   100.0%;  Pred. No. 8.6e-105;
  Matches  270;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MELNRSEADEAKAETTPTGGATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDSPNV 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 melnrseadeakaettptggatssatasgsssgrrprgrpagsknkpkpptiitrdspnv 60

Qy      61 LRSHVLEVTSGSDISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGR 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 lrshvlevtsgsdiseavstyatrrgcgvciisgtgavtnvtirqpaapagggvitlhgr 120

Qy     121 FDILSLTGTALPPPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVY 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 fdilsltgtalpppappgaggltvylaggqgqvvggnvagsliasgpvvlmaasfanavy 180

Qy     181 DRLPIEEEETPPPRTTGVQQQQPEASQSSEVTGSGAQACESNLQGGNGGGGVAFYNLGMN 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 drlpieeeetppprttgvqqqqpeasqssevtgsgaqacesnlqggngggggvafynlgmn 240

Qy     241 MNNFQFSGGDIYGMSGGSGGGGGGATRPAF 270
           |||||||||||||||||||||||||||||
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2002, 21:20:07 ; Search time 69.12 Seconds
(without alignments)
433.882 Million cell updates/sec

Title: US-09-823-676-2
Perfect score: 1392
Sequence: 1 MELNRSEADEAKAETTPTGG.....IYGMSGGSGGGGGGATRPAF 270

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802:*
1: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1392	100.0	270	22	AAE01961	Arabidopsis thalia
2	1392	100.0	270	22	AAB67838	Amino acid sequenc
3	646.5	46.4	254	21	AAG29345	Arabidopsis thalia
4	646.5	46.4	292	21	AAG29344	Arabidopsis thalia
5	637.5	45.8	280	21	AAG10650	Arabidopsis thalia
6	637.5	45.8	282	21	AAG10649	Arabidopsis thalia
7	637.5	45.8	315	21	AAG10648	Arabidopsis thalia
8	625.5	44.9	339	22	AAE02571	A. thaliana transc
9	572	41.1	272	21	AAG21497	Arabidopsis thalia
10	572	41.1	272	21	AAG54092	Arabidopsis thalia
11	572	41.1	310	21	AAG21496	Arabidopsis thalia
12	572	41.1	310	21	AAG54091	Arabidopsis thalia
13	572	41.1	315	21	AAG54090	Arabidopsis thalia
14	572	41.1	316	21	AAG21495	Arabidopsis thalia
15	570	40.9	239	21	AAG41483	Arabidopsis thalia
16	570	40.9	239	21	AAG45262	Arabidopsis thalia
17	570	40.9	267	21	AAG41482	Arabidopsis thalia
18	570	40.9	267	21	AAG45261	Arabidopsis thalia
19	570	40.9	281	21	AAG41481	Arabidopsis thalia
20	570	40.9	281	21	AAG45260	Arabidopsis thalia
21	274	19.7	336	21	AAG46021	Arabidopsis thalia
22	274	19.7	343	21	AAG46020	Arabidopsis thalia
23	274	19.7	351	21	AAG46019	Arabidopsis thalia
24	255.5	18.4	304	21	AAG20509	Arabidopsis thalia
25	255.5	18.4	304	21	AAG35061	Arabidopsis thalia
26	255.5	18.4	334	21	AAG20508	Arabidopsis thalia
27	255.5	18.4	334	21	AAG35060	Arabidopsis thalia
28	255.5	18.4	334	22	AAE01895	Arabidopsis thalia
29	249	17.9	272	21	AAG20510	Arabidopsis thalia
30	249	17.9	272	21	AAG35062	Arabidopsis thalia
31	248.5	17.9	386	22	AAE02570	A. thaliana transc
32	243	17.5	386	22	AAE02551	A. thaliana transc
33	234.5	16.8	356	22	AAE01908	Arabidopsis thalia
34	233.5	16.8	277	21	AAG10705	Arabidopsis thalia
35	233.5	16.8	352	21	AAG10704	Arabidopsis thalia
36	233.5	16.8	356	21	AAG10703	Arabidopsis thalia
37	227	16.3	290	21	AAG34779	Arabidopsis thalia
38	227	16.3	365	21	AAG34778	Arabidopsis thalia
39	227	16.3	369	21	AAG34777	Arabidopsis thalia
40	225	16.2	264	21	AAG22421	Arabidopsis thalia
41	225	16.2	264	21	AAG40514	Arabidopsis thalia
42	225	16.2	342	21	AAG22420	Arabidopsis thalia
43	225	16.2	342	21	AAG40513	Arabidopsis thalia
44	225	16.2	348	21	AAG22419	Arabidopsis thalia
45	225	16.2	348	21	AAG40512	Arabidopsis thalia

ALIGNMENTS

RESULT 1
AAE01961
ID AAE01961 standard; Protein; 270 AA.
XX
AC AAE01961;
XX
DT 31-JUL-2001 (first entry)
XX
DE Arabidopsis thaliana transcription factor G1073.
XX
KW Transcription factor; trait modification; structural characteristic;
KW developmental characteristic; gene therapy; agricultural biotechnology.
XX
OS Arabidopsis thaliana.
XX
FH Key Location/Qualifiers
FT Domain 33..42
FT /note= "Conserved domain"
FT Domain 78..175
FT /note= "Conserved domain"
XX
PN WO200136444-A1.
XX
PD 25-MAY-2001.
XX
PF 14-NOV-2000; 2000WO-US31325.
XX
PR 17-NOV-1999; 99US-0166228.
PR 17-APR-2000; 2000US-0197899.
PR 22-AUG-2000; 2000US-0227439.
XX
PA (MEND-) MENDEL BIOTECHNOLOGY INC.
PA (RIEC/) RIECHMANN J L.
PA (REUB/) REUBER L.

```
FT   CARBOHYD      81     81       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD      83     83       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD      87     87       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD      91     91       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD      93     93       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD      94     94       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD      96     96       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD      98     98       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     101    101       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     103    103       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     104    104       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     106    106       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     107    107       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     108    108       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     110    110       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     114    114       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     117    117       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     123    123       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     124    124       O-LINKED (GALNAC. . .) (MUCIN TYPE).
FT   CARBOHYD     418    418       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     547    547       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     917    917       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     985    985       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1002   1002       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1068   1068       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1150 AA;  109615 MW;  3CB68B5D29DD7F5A CRC64;
```

```
  Query Match             9.2%;  Score 128.5;  DB 1;  Length 1150;
  Best Local Similarity   26.6%;  Pred. No. 0.47;
  Matches   79;  Conservative   30;  Mismatches  107;  Indels   81;  Gaps   17;

Qy      18 TGGATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDSPNVLRSHVLEVTSGSDISEA 77
           ||  : :: ::|||||    |    |  :|:   |    |     |    |:||:  |:|
Db      13 TGTVSGASGSTGSSSG--STGATGASIGQPETSRISVAGS-----SGAPAVSSGA--SQA 63

Qy      78 VSTYATRRGCGVCIISGTG---AVTNVTIRQPAAP--AGGGVI-TLHGRFDILSLTGTAL 131
             |           || |     ::| : : | |  || |   |: |       :|:
Db      64 AGT-----------SGAGPGTTASSVGVTETARPSVAGSGTTGTVSG------ASGSTG 105

Qy     132 PPPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVYDRLP------- 184
               ||| | ::    || :    :|||| :  | |   || |       |
Db     106 SSSGSPGATGASI----GQPETSRISVAGS--SGAPAVSSGASQAAGTSGAGPGTTASSV 159

Qy     185 -IEEEETP----------------------PPRTTGVQQQQPEASQSSEVTGSGAQACE 220
            : |   |                      |  ||    ||| |: |    ||| |
Db     160 GVTETARPSVAGSGTTGTVSGASGSTGSSSGSPGATGASIGQPETSRISVAGSSGAPAVS 219

Qy     221 SNLQ---GGNGGG-GVAFYNLGMNMN---NFQFSG--GDIYGMSG--GSGGGGGGAT 266
           |      | :| | |    ::|:      :   ||  | : | ||  ||  |  |||
Db     220 SGASQAAGTSGAGPGTTASSVGVTETARPSVAGSGTTGTVSGASGSTGSSSGSPGAT 276
```

```
RESULT   3
FSH_DROME
ID   FSH_DROME      STANDARD;      PRT;  2038 AA.
AC   P13709; P13710;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   Female sterile homeotic protein (Fragile-chorion membrane protein).
GN   FS(1)H OR FSH.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89276730; PubMed=2567251;
RA   Haynes S.R., Mozer B.A., Bhatia-Dey N., Dawid I.B.;
RT   "The Drosophila fsh locus, a maternal effect homeotic gene, encodes
RT   apparent membrane proteins.";
RL   Dev. Biol. 134:246-257(1989).
CC   -!- FUNCTION: REQUIRED MATERNALLY FOR PROPER EXPRESSION OF OTHER
CC       HOMEOTIC GENES INVOLVED IN PATTERN FORMATION, SUCH AS UBX.
CC   -!- SIMILARITY: HIGH, TO HUMAN RING3 PROTEIN.
CC   -!- SIMILARITY: CONTAINS 2 BROMODOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 ET DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M23221; AAA28540.1; -.
DR   EMBL; M23222; AAA28541.1; ALT_TERM.
DR   EMBL; M15762; AAA70424.1; -.
DR   EMBL; M15763; AAA70423.1; -.
DR   EMBL; M15764; AAA70422.1; -.
DR   PIR; A43742; A43742.
DR   HSSP; P04002; 1WFA.
DR   FlyBase; FBgn0004656; fs(1)h.
DR   InterPro; IPR001487; Bromodomain.
DR   Pfam; PF00439; bromodomain; 2.
DR   PRINTS; PR00503; BROMODOMAIN.
DR   SMART; SM00297; BROMO; 2.
DR   PROSITE; PS00633; BROMODOMAIN_1; 2.
DR   PROSITE; PS50014; BROMODOMAIN_2; 2.
KW   Developmental protein; Bromodomain; Transmembrane; Repeat.
FT   DOMAIN       51    123       BROMODOMAIN 1.
FT   DOMAIN      495    567       BROMODOMAIN 2.
FT   DOMAIN      945   1106       ET DOMAIN.
FT   TRANSMEM    330    350       POTENTIAL.
FT   TRANSMEM    451    471       POTENTIAL.
FT   TRANSMEM    750    770       POTENTIAL.
FT   TRANSMEM    790    810       POTENTIAL.
FT   TRANSMEM    816    830       POTENTIAL.
FT   TRANSMEM    874    894       POTENTIAL.
FT   TRANSMEM   1731   1751       POTENTIAL.
FT   TRANSMEM   1939   1959       POTENTIAL.
FT   VARIANT     909    909       G -> A.
FT   VARIANT    1022   1022       H -> RKPYY.
SQ   SEQUENCE   2038 AA;  205332 MW;  849E0706D50A0098 CRC64;
```

```
  Query Match             9.0%;  Score 125.5;  DB 1;  Length 2038;
  Best Local Similarity   23.1%;  Pred. No. 1.3;
  Matches   73;  Conservative   31;  Mismatches  111;  Indels  101;  Gaps   13;

Qy       7 EADEAKAETTPTGGATSSA------TASGSSSGRRPRGRPAGS-----KNKPKPPTIITR 55
           :|:|   :   :  : :||      |:| |||    : :||       :|| |      |
Db    1633 KANEQNLKNASSWSSLASANSPQSHTSSSSSSS---KAKPAMDSFQQFRNKAK-----ER 1684

Qy      56 DSPNVLRSHVLEVTSGSDISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPA---GG 112
           |   :| :   |  :  : :|       :       | : ::|:  : | || |
Db    1685 DRLKLLEAAEKEKKNQKEAAEKEQQRKHHK-------SSSSSLTSAAVAQAAAIAAATAA 1737

Qy     113 GVITLHGRFDILSLTGTALPPPAPPGAGGLTVYLAGGQGQVV----------------- 154
             :|| |     :|  :|  |     :||     ||   |
Db    1738 AAVTL-GAAAAAALASSASNPSGGSSSGG-----AGSTSQQAITGDRDRDRDRERERERS 1791

Qy     155 ---------GGNVAGSLIASGPVVLMAASFANAVYDRLPIEEEETPPPRTTGVQQQQPEA 205
                   | | : |  ::||    :                          :|      |
Db    1792 GSGGGQSGNGNNSSNSANSNGPGSAGSGG--------------------SGGGGGSGPA 1830

Qy     206 SQSSEVTGSGAQACESNLQGGNGGGGVAFYNLGMNMN----------NFQFSGGDIYGM 254
           |      :| |  |  ||  || |||| |  | | | |          |   | | | |
Db    1831 SAGGPNSGGGGTA-NSNSGGGGGGGGPALLNAGSNSNSGVGSGGAASSNSNSSVGGIVGS 1889

Qy     255 SG------GSGGGGGG 264
             |      || |||||
Db    1890 GGPGSNSQGSSGGGGG 1905
```

```
RESULT   4
MEF2_DROME
ID   MEF2_DROME     STANDARD;      PRT;   515 AA.
AC   P40791;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Myocyte-specific enhancer factor 2.
GN   MEF2.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RX   MEDLINE=94261646; PubMed=8202544;
RA   Lilly B., Galewsky S., Firulli A.B., Schulz R.A., Olson E.N.;
RT   "D-MEF2: a MADS box transcription factor expressed in differentiating
RT   mesoderm and muscle cell lineages during Drosophila embryogenesis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 91:5662-5666(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94329547; PubMed=8052612;
RA   Nguyen H.T., Bodmer R., Abmayr S.M., McDermott J.C., Spoerel N.A.;
RT   "D-mef2: a Drosophila mesoderm-specific MADS box-containing gene with
RT   a biphasic expression profile during embryogenesis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 91:7520-7524(1994).
CC   -!- FUNCTION: TRANSCRIPTION FACTOR THAT COULD BE A KEY PLAYER IN EARLY
CC       MESODERM DIFFERENTIATION AND MAY BE REQUIRED FOR SUBSEQUENT CELL
CC       FATE SPECIFICATIONS WITHIN THE SOMATIC AND VISCERAL/HEART
CC       MESODERMAL LAYERS.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- TISSUE SPECIFICITY: MESODERM.
CC   -!- DEVELOPMENTAL STAGE: FIRST DETECTABLE IN THE PRESUMPTIVE MESODERM
CC       AT LATE CELLULAR BLASTODERM STAGE. EXPRESSED IN ALL PRESUMPTIVE
CC       MESODERM PRIOR TO THE SPLITTING PROCESS THAT GENERATES THE SOMATIC
CC       AND VISCERAL/ HEART MESODERM. AFTER THE SUBDIVISION, IT IS FOUND
CC       IN BOTH THE SOMATIC AND THE VISCERAL/HEART MESODERM.
CC   -!- INDUCTION: TWI ACTIVITY IS REQUIRED FOR MEF2 EXPRESSION. SNA
CC       ACTIVITY IS NEEDED FOR MAINTAINING MEF2 EXPRESSION.
CC   -!- SIMILARITY: BELONGS TO THE MADS DOMAIN FAMILY OF TRANSCRIPTION
CC       FACTORS. MEF2 SUBFAMILY.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; U03292; AAA19957.1; -.
DR   EMBL; U07422; AAA20463.1; -.
DR   HSSP; P11831; 1SRS.
DR   TRANSFAC; T01772; -.
DR   FlyBase; FBgn0011656; Mef2.
DR   InterPro; IPR002100; MADS-box.
DR   Pfam; PF00319; SRF-TF; 1.
DR   PRINTS; PR00404; MADSDOMAIN.
DR   SMART; SM00432; MADS; 1.
DR   PROSITE; PS00350; MADS_BOX_1; 1.
DR   PROSITE; PS50066; MADS_BOX_2; 1.
KW   Transcription regulation; DNA-binding; Nuclear protein;
KW   Developmental protein; Differentiation; Activator.
FT   DOMAIN        3     57       MADS.
FT   DNA_BIND     58     86       MEF2-TYPE (POTENTIAL).
FT   DOMAIN      380    386       POLY-SER.
FT   DOMAIN      390    396       POLY-GLY.
FT   DOMAIN      421    428       POLY-GLY.
FT   DOMAIN      484    495       POLY-GLN.
FT   CONFLICT    338    338       S -> R (IN REF. 2).
FT   CONFLICT    364    377       PAVISRIAMVPRAG -> ASGHQQNSNGSTGS (IN
FT                                REF. 2).
SQ   SEQUENCE   515 AA;  54289 MW;  832C1FB4A831F28E CRC64;


  Query Match             8.9%;  Score 123.5;  DB 1;  Length 515;
  Best Local Similarity   23.1%;  Pred. No. 0.43;
  Matches   66;  Conservative   31;  Mismatches  114;  Indels   75;  Gaps   12;

Qy      16 TPTGGATSSATASGS----SSGRRPRGRPAGSKNKPKPPTIITRDSPNVLRSHVLEVTSG 71
           | :|| :     |||    |:|      |      | |       || :  :| | :
Db     184 TDSGGMSLIIYPSGSMLEMSNGYPHSHSPLVGSPSPGP-------SPGI--AHHLSIKQQ 234

Qy      72 SDISEAVSTYATRRGCGVCIISGTGAVTNVTIRQP----------AAPAGGGVITLHGRF 121
           |  |:                    |  :|: :  |          :||   |
Db     235 SPGSQ-----------------NGRASNLRVVIPPTIAPIPPNMSAPDDVGYADQRQSQ 276

Qy     122 DILSLTGTALPPPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVYD 181
             |:     |  | |    || |  | |      : : | : :   ::   ::   :
Db     277 TSLNTPVVTLQTPIP----ALTSYSFGAQ------DFSSSGVMNSADIMSLNTWHQGLVP 326

Qy     182 RLPIE----EEETPPPRTTGV------QQQQPEASQSSEVT----------GSGAQACES 221
             :        |||| |: |       : | |    |  |                |  : :  |
Db     327 HSSLSHLAVSNSTPPPATSPVSIKVKAEPQSPPRDLSPAVISRIAMVPRAGGGSSSSTSS 386

Qy     222 NLQGGNGGGG-VAFYNLGMNMNNFQFSGGDIYGMSGGSGGGGGGAT 266
           |  || |||| |:  |:  ::||     |   | ||  ||||||  :
Db     387 NASGGAGGGGAVSAANVITHLNNVSVLAG---GPSGQGGGGGGGGS 429


RESULT   5
ANDR_EULFC
ID   ANDR_EULFC     STANDARD;      PRT;   884 AA.
AC   O97776;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Androgen receptor (Dihydrotestosterone receptor).
GN   AR OR NR3C4.
OS   Eulemur fulvus collaris (Collared brown lemur).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Strepsirhini; Lemuridae; Eulemur.
OX   NCBI_TaxID=47178;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98404153; PubMed=9732460;
RA   Choong C.S., Kemppainen J.A., Wilson E.M.;
RT   "Evolution of the primate androgen receptor: a structural basis for
RT   disease.";
RL   J. Mol. Evol. 47:334-342(1998).
CC   -!- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN
CC       THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR
CC       PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC       A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
CC       NR3 SUBFAMILY.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; U94178; AAC73049.1; -.
DR   HSSP; P06536; 1GDC.
DR   InterPro; IPR001103; Androgen_recep.
DR   InterPro; IPR000536; Hormone_rec_lig.
```

```
DR   InterPro; IPR001628; zf-C4.
DR   Pfam; PF02166; Androgen_recep; 1.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00047; STROIDFINGER.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW   Zinc-finger; Steroid-binding.
FT   DOMAIN        1    522       MODULATING (BY SIMILARITY).
FT   DNA_BIND    524    589       NUCLEAR RECEPTOR-TYPE.
FT   ZN_FING     524    544       C4-TYPE.
FT   ZN_FING     560    584       C4-TYPE.
FT   DOMAIN      655    884       LIGAND-BINDING.
FT   DOMAIN       55     58       POLY-GLN.
FT   DOMAIN       64     70       POLY-GLN.
FT   DOMAIN      116    120       POLY-ALA.
FT   DOMAIN      174    178       POLY-GLN.
FT   DOMAIN      353    362       POLY-PRO.
FT   DOMAIN      379    383       POLY-ALA.
FT   DOMAIN      408    411       POLY-ALA.
FT   DOMAIN      430    435       POLY-GLY.
SQ   SEQUENCE   884 AA;  95610 MW;  18F570E352F4D2BD CRC64;


  Query Match             8.9%;  Score 123.5;  DB 1;  Length 884;
  Best Local Similarity   25.1%;  Pred. No. 0.74;
  Matches   76;  Conservative   28;  Mismatches  136;  Indels   63;  Gaps   12;

Qy       1 MELNRSEADEAKAETTPTGGATSSATASGSSSG--RRPRGRPAGSKNKPKPPTIITRDSP 58
           |:| : :  || :|              |||||  |   | |  ||:     |    ||
Db     170 MQLLQQQQQEAVSE-------------GSSSGRAREAAGAPTSSKDSYLGGTSTISDSA 215

Qy      59 NVLRSHV-LEVTSGSDISEAVSTYATRRG-CGVCIISGTGAVTNVTIRQPAAPAGGGVIT 116
             |   | : :  | :  | :|     || |    : |       | | |   | ::
Db     216 KELCKAVSVSMGLGVETLEHLSPGEQLRGDCMYAPLLGGPPAVRPTPCAPLAECKGSLL- 274

Qy     117 LHGRFDILSLTGTALPPPAPPGAGGLTVYLAGGQGQVVGGNVAG----------SLIAS 165
                | :  ||  |    |  | |  | | |      | : ||           || |
Db     275 -----DDSADKGTEEPAEYTPFKGSYTQGLEGESLGCSGSSEAGSSGTLELPSTLSLYKS 329

Qy     166 GPVVLMAASFANAVYDRLPIEEEETPPPRT-----TGVQQQQPEASQSSEVTGSGAQACE 220
           |  :  |||: :  |   |:     |||         :: : |      |    : ||
Db     330 G-ALEEAASYQSRDYYNFPLALAGPPPPPLPPHPHARIKLENP-LDYGSSWAAAAAQCRF 387

Qy     221 SNLQGGNGGGGV------------AFYNLGMNMNNFQFSGGDIYGMSGGSGGG--GGGA 265
           :|   :|||              ::: |      |    | :||  || |||    ||
Db     388 GDLASLHGGGATGPGSGSPSAAAASSWHTL------FTAEEGQLYGPCGGGGGGTSEAGA 441

Qy     266 TRP 268
           |
Db     442 VTP 444


RESULT   6
ANDR_MACFA
ID   ANDR_MACFA     STANDARD;      PRT;   895 AA.
AC   O97952;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Androgen receptor (Dihydrotestosterone receptor).
GN   AR OR NR3C4.
OS   Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9541;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98404153; PubMed=9732460;
RA   Choong C.S., Kemppainen J.A., Wilson E.M.;
RT   "Evolution of the primate androgen receptor: a structural basis for
RT   disease.";
RL   J. Mol. Evol. 47:334-342(1998).
CC   -!- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN
CC       THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR
CC       PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC       A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
CC       NR3 SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U94179; AAC73050.1; -.
DR   HSSP; P06536; 1GDC.
DR   InterPro; IPR001103; Androgen_recep.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001628; zf-C4.
DR   Pfam; PF02166; Androgen_recep; 1.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00047; STROIDFINGER.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW   Zinc-finger; Steroid-binding.
FT   DOMAIN        1    533       MODULATING (BY SIMILARITY).
FT   DNA_BIND    535    600       NUCLEAR RECEPTOR-TYPE.
FT   ZN_FING     535    555       C4-TYPE.
FT   ZN_FING     571    595       C4-TYPE.
FT   DOMAIN      666    895       LIGAND-BINDING.
FT   DOMAIN       55     62       POLY-GLN.
FT   DOMAIN       68     74       POLY-GLN.
FT   DOMAIN      178    182       POLY-GLN.
FT   DOMAIN      357    366       POLY-PRO.
FT   DOMAIN      381    387       POLY-ALA.
FT   DOMAIN      434    448       POLY-GLY.
SQ   SEQUENCE   895 AA;  96494 MW;  A3EB17916F43A097 CRC64;


  Query Match             8.8%;  Score 122.5;  DB 1;  Length 895;
  Best Local Similarity   23.4%;  Pred. No. 0.86;
  Matches   70;  Conservative   35;  Mismatches  129;  Indels   65;  Gaps   11;

Qy       1 MELNRSEADEAKAETTPTGGATSSATASGSSSG--RRPRGRPAGSKNKPKPPTIITRDSP 58
           |:| : :  || :|              |||||  |   | |  ||:     |    ||
Db     174 MQLLQQQQQEAVSE-------------GSSSGRAREASGAPTSSKDNYLGGTSTISDSA 219

Qy      59 NVLRSHV-LEVTSGSDISEAVSTYATRRG-CGVCIISGTGAVTNVTIRQPAAPAGGGVIT 116
             |   | : :  | :  | :|     || |    : |       | | |   | ::
Db     220 KELCKAVSVSMGLGVEALEHLSPGEQLRGDCMYAPVLGVPPAVRPTPCAPLAECKGSLL- 278

Qy     117 LHGRFDILSLTGTALPPPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVL------ 170
                |  :  |       |  ||   |  | :|: :| : : :  :|| : |
Db     279 -----DDSAGKSTEDTAEYSPFKGG---YTKGLEGESLGCSGSAAAGSSGTLELPSTLSL 330

Qy     171 -------MAASFANAVYDRLPIEEEETPPPRT-----TGVQQQQPEASQSSEVTGSGAQA 218
                 ||:: :  |   |:     |||         :: : |      |    : ||
Db     331 YKSGALDEAAAYQSRDYYNFPLALAGPPPPPPPPHPHARIKLENP-LDYGSAWAAAAAQC 389

Qy     219 CESNLQGGNGGGGV------------AFYNLGMNMNNFQFSGGDIYGMSGGSGGGGGG 264
           :|   :| |                ::: |      |    | :||  || ||||||
Db     390 RYGDLASLHGAGAAGPGSGSPSAAASSSWHTL------FTAEEGQLYGPCGGGGGGGG 442
```

```
RESULT   7
RFX1_HUMAN
ID   RFX1_HUMAN     STANDARD;      PRT;   979 AA.
AC   P22670;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   MHC class II regulatory factor RFX1 (RFX) (Enhancer factor C) (EF-C).
GN   RFX1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91071581; PubMed=2253877;
RA   Reith W., Sanchez-Herrero C., Kobr M., Silacci P., Berte C.,
RA   Barras E., Mach B.;
RT   "MHC class II regulatory factor RFX has a novel DNA-binding domain
RT   and a functionally independent dimerization domain.";
RL   Genes Dev. 4:1528-1540(1990).
RN   [2]
RP   IDENTITY BETWEEN RFX1 AND EF-C.
RX   MEDLINE=94019311; PubMed=8413236;
RA   Siegrist C.A., Durand B., Emery P., David E., Hearing P., Mach B.,
RA   Reith W.;
RT   "RFX1 is identical to enhancer factor C and functions as a
RT   transactivator of the hepatitis B virus enhancer.";
RL   Mol. Cell. Biol. 13:6375-6384(1993).
RN   [3]
RP   BINDING TO RPL30 PROMOTER.
RX   MEDLINE=94040774; PubMed=8224874;
RA   Safrany G., Perry R.P.;
RT   "Transcription factor RFX1 helps control the promoter of the mouse
RT   ribosomal protein-encoding gene rpL30 by binding to its alpha
RT   element.";
RL   Gene 132:279-283(1993).
RN   [4]
RP   SHOWS THAT BLS IS NOT DUE TO RFX1.
RX   MEDLINE=92375076; PubMed=1508204;
RA   Sanchez-Herrero C., Reith W., Silacci P., Mach B.;
RT   "The DNA-binding defect observed in major histocompatibility complex
RT   class II regulatory mutants concerns only one member of a family of
RT   complexes binding to the X boxes of class II promoters.";
RL   Mol. Cell. Biol. 12:4076-4083(1992).
CC   -!- FUNCTION: REGULATORY FACTOR ESSENTIAL FOR MHC CLASS II GENE
CC       EXPRESSION. BINDS TO THE X BOXES OF MHC CLASS II GENES. ALSO BINDS
CC       TO AN INVERTED REPEAT (ENH1) REQUIRED FOR HEPATITIS B VIRUS GENES
CC       EXPRESSION AND TO THE MOST UPSTREAM ELEMENT (ALPHA) OF THE RPL30
CC       PROMOTER.
CC   -!- SUBUNIT: BINDS DNA AS AN HOMODIMER.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: BELONGS TO THE RFX FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X58964; CAA41730.1; -.
DR   EMBL; A20498; CAA01506.1; -.
DR   PIR; A35913; A35913.
DR   TRANSFAC; T00909; -.
DR   TRANSFAC; T01673; -.
DR   MIM; 600006; -.
DR   InterPro; IPR003150; RFX_DNA_binding.
DR   Pfam; PF02257; RFX_DNA_binding; 1.
KW   DNA-binding; Transcription regulation; Activator; Nuclear protein.
FT   DOMAIN      381    411       GLY-RICH.
FT   DNA_BIND    438    528       EXPERIMENTALLY DEDUCED.
FT   DOMAIN      920    936       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN      744    979       NECESSARY FOR DIMERIZATION.
SQ   SEQUENCE   979 AA;  104728 MW;  556151F88C6AC9A2 CRC64;


  Query Match             8.8%;  Score 122;  DB 1;  Length 979;
  Best Local Similarity   20.5%;  Pred. No. 1;
  Matches   78;  Conservative   34;  Mismatches  124;  Indels  144;  Gaps   14;

Qy      17 PTGGATSSATASGSSSGRRPRGRPAGSKNK----------PKPPTIITRDSPNVLRSHVL 66
           ||  ||            :|: :| | : :          |  ||     ||   :  |:
Db      44 PTAAATPQPQYVTELQSPQPQAQPPGGQKQYVTELPAVPAPSQPTGAPTPSPAPQQYIVV 103

Qy      67 EVTSGS-DISEAVS-----TYATRRGCGVCI---ISGTG------------------AV 98
            |: |:   || ||      : |:: |    :   : ||                   :
Db     104 TVSEGAMRASETVSEASPGSTASQTGVPTQVVQQVQGTQQRLLVQTSVQAKPGHVSPLQL 163

Qy      99 TNVTIRQPAAPA-------------GGGV-ITLHGRFDILS------------------- 125
           ||: : | | |                || | :|:||   : |
Db     164 TNIQVPQQALPTQRLVVQSAAPGSKGGQVSLTVHGTQQVHSPPEQSPVQANSSSSKTAGA 223

Qy     126 LTGT-----------------------ALPPPAPPG-AGGLTV---------------- 144
            |||                       | |    ||    |||
Db     224 PTGTVPQQLQVHGVQQSVPVTQERSVVQATPQAPKPGPVQPLTVQGLQPVHVAQEVQQLQ 283

Qy     145 -------------YLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVYDRLPIEEEETP 191
                       |: ||        :  | :    |   : : : |:      : :
Db     284 QVPVPHVYSSQVQYVEGGDASYTASAIRSSTYSYPETPLYTQTASTSYYEAAGTATQVST 343

Qy     192 PPRTTGVQQ--QQPEASQSSEVTGS----GAQACESNLQGGNGGGGVAFYNLGMNMNNFQ 245
           |  :  |      |      |:|  |    || |  |:  ||:||||
Db     344 PATSQAVASSGSMPMYVSGSQVVASSASTGAGASNSSGGGGSGGGGGG------------ 391

Qy     246 FSGGDIYGMSGGSGGGGGGA 265
             ||   | || :|||| ||
Db     392 -GGGGGGGGSGSTGGGGSGA 410


RESULT   8
ANDR_PAPHA
ID   ANDR_PAPHA     STANDARD;      PRT;   895 AA.
AC   O97960;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Androgen receptor (Dihydrotestosterone receptor).
GN   AR OR NR3C4.
OS   Papio hamadryas (Hamadryas baboon).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Papio.
OX   NCBI_TaxID=9557;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98404153; PubMed=9732460;
RA   Choong C.S., Kemppainen J.A., Wilson E.M.;
RT   "Evolution of the primate androgen receptor: a structural basis for
RT   disease.";
RL   J. Mol. Evol. 47:334-342(1998).
CC   -!- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN
CC       THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR
CC       PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC       A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
CC       NR3 SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
```

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; U94176; AAC73047.1; -.
DR   HSSP; P06536; 1GDC.
DR   InterPro; IPR001103; Androgen_recep.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001628; zf-C4.
DR   Pfam; PF02166; Androgen_recep; 1.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00047; STROIDFINGER.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW   Zinc-finger; Steroid-binding.
FT   DOMAIN        1    533       MODULATING (BY SIMILARITY).
FT   DNA_BIND    535    600       NUCLEAR RECEPTOR-TYPE.
FT   ZN_FING     535    555       C4-TYPE.
FT   ZN_FING     571    595       C4-TYPE.
FT   DOMAIN      666    895       LIGAND-BINDING.
FT   DOMAIN       55     63       POLY-GLN.
FT   DOMAIN       69     74       POLY-GLN.
FT   DOMAIN      178    182       POLY-GLN.
FT   DOMAIN      357    366       POLY-PRO.
FT   DOMAIN      381    387       POLY-ALA.
FT   DOMAIN      434    448       POLY-GLY.
SQ   SEQUENCE   895 AA;  96478 MW;  9020C0DC67F11E5D CRC64;


  Query Match             8.7%;  Score 121.5;  DB 1;  Length 895;
  Best Local Similarity   23.4%;  Pred. No. 0.99;
  Matches   70;  Conservative   34;  Mismatches  130;  Indels   65;  Gaps   11;

Qy       1 MELNRSEADEAKAETTPTGGATSSATASGSSSG--RRPRGRPAGSKNKPKPPTIITRDSP 58
           |:| : :  || :|              |||||  |   | |  ||:     |    ||
Db     174 MQLLQQQQQEAVSE-------------GSSSGRAREASGAPTSSKDNYLGGTSTISDSA 219

Qy      59 NVLRSHV-LEVTSGSDISEAVSTYATRRG-CGVCIISGTGAVTNVTIRQPAAPAGGGVIT 116
             |   | : :  | :  | :|     || |    : |       |   | |   | ::
Db     220 KELCKAVSVSMGLGVEALEHLSPGEQLRGDCMYAPVLGVPPAVRPTPCAPLAECKGSLL- 278

Qy     117 LHGRFDILSLTGTALPPPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVL------ 170
               |  :   |       |  ||   |  | :|: :| : : :  :|| : |
Db     279 -----DDSAGKSTEDTAEYSPFKGG---YTKGLEGESLGCSGSAAAGSSGTLELPSTLSL 330

Qy     171 -------MAASFANAVYDRLPIEEEETPPPRT-----TGVQQQQPEASQSSEVTGSGAQA 218
                 ||:: :  |   |:     |||         :: : |     |    : ||
Db     331 YKSGALDEAAAYQSRDYYNFPLALAGPPPPPPPPHPHARIKLENP-LDYGSAWAAAAAQC 389

Qy     219 CESNLQGGNGGGGV-------------AFYNLGMNMNNFQFSGGDIYGMSGGSGGGGGG 264
             |   :| |                 ::: |      |    | :||  || ||||||
Db     390 RYGELASLHGAGAAGPGSGSPSAAASSSWHTL------FTAEEGQLYGPCGGGGGGGG 442


RESULT   9
ANDR_PANTR
ID   ANDR_PANTR     STANDARD;      PRT;   911 AA.
AC   O97775;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Androgen receptor (Dihydrotestosterone receptor).
GN   AR OR NR3C4.
OS   Pan troglodytes (Chimpanzee).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX   NCBI_TaxID=9598;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98404153; PubMed=9732460;
RA   Choong C.S., Kemppainen J.A., Wilson E.M.;
RT   "Evolution of the primate androgen receptor: a structural basis for
RT   disease.";
RL   J. Mol. Evol. 47:334-342(1998).
CC   -!- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN
CC       THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR
CC       PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC       A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
CC       NR3 SUBFAMILY.
CC   ------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; U94177; AAC73048.1; -.
DR   HSSP; P06536; 1GDC.
DR   InterPro; IPR001103; Androgen_recep.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001628; zf-C4.
DR   Pfam; PF02166; Androgen_recep; 1.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00047; STROIDFINGER.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW   Zinc-finger; Steroid-binding.
FT   DOMAIN        1    549       MODULATING (BY SIMILARITY).
FT   DNA_BIND    551    616       NUCLEAR RECEPTOR-TYPE.
FT   ZN_FING     551    571       C4-TYPE.
FT   ZN_FING     587    611       C4-TYPE.
FT   DOMAIN      682    911       LIGAND-BINDING.
FT   DOMAIN       57     78       POLY-GLN.
FT   DOMAIN       84     88       POLY-GLN.
FT   DOMAIN      192    196       POLY-GLN.
FT   DOMAIN      371    380       POLY-PRO.
FT   DOMAIN      395    401       POLY-ALA.
FT   DOMAIN      448    464       POLY-GLY.
SQ   SEQUENCE   911 AA;  98402 MW;  601B9BD4E697DAA4 CRC64;


  Query Match             8.7%;  Score 121.5;  DB 1;  Length 911;
  Best Local Similarity   23.4%;  Pred. No. 1;
  Matches   70;  Conservative   35;  Mismatches  129;  Indels   65;  Gaps   11;

Qy       1 MELNRSEADEAKAETTPTGGATSSATASGSSSG--RRPRGRPAGSKNKPKPPTIITRDSP 58
           |:| : :  || :|              |||||  |   | |  ||:     |    ||
Db     188 MQLLQQQQQEAVSE-------------GSSSGRAREASGAPTSSKDNYLGGTSTISDSA 233

Qy      59 NVLRSHV-LEVTSGSDISEAVSTYATRRG-CGVCIISGTGAVTNVTIRQPAAPAGGGVIT 116
             |   | : :  | :  | :|     || |    : |       |   | |   | ::
Db     234 KELCKAVSVSMGLGVEALEHLSPGEQLRGDCMYAPLLGVPPAVRPTPCAPLAECKGSLL- 292

Qy     117 LHGRFDILSLTGTALPPPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVL------ 170
               |  :   |       |  ||   |  | :|: :| : : :  :|| : |
Db     293 -----DDSAGKSTEDTAEYSPFKGG---YTKGLEGESLGCSGSAAAGSSGTLELPSTLSL 344

Qy     171 -------MAASFANAVYDRLPIEEEETPPPRT-----TGVQQQQPEASQSSEVTGSGAQA 218
                 ||:: :  |   |:     |||         :: : |     |    : ||
Db     345 YKSGALDEAAAYQSRDYYNFPLALAGPPPPPPPPHPHARIKLENP-LDYGSAWAAAAAQC 403

Qy     219 CESNLQGGNGGGGV-------------AFYNLGMNMNNFQFSGGDIYGMSGGSGGGGGG 264
             :|   :| |                 ::: |      |    | :||  || ||||||
Db     404 RYGDLASLHGAGAAGPGSGSPSAAASSSWHTL------FTAEEGQLYGPCGGGGGGGG 456
```

```
RESULT  10
ANDR_HUMAN
ID   ANDR_HUMAN     STANDARD;      PRT;   919 AA.
AC   P10275;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Androgen receptor (Dihydrotestosterone receptor).
GN   AR OR NR3C4 OR DHTR.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89112208; PubMed=3216866;
RA   Lubahn D.B., Joseph D.R., Sar M., Tan J., Higgs H.N., Larson R.E.,
RA   French F.S., Wilson E.M.;
RT   "The human androgen receptor: complementary deoxyribonucleic acid
RT   cloning, sequence analysis and gene expression in prostate.";
RL   Mol. Endocrinol. 2:1265-1275(1988).
RN   [2]
RP   SEQUENCE FROM N.A., AND VARIANT CAIS MET-866.
RX   MEDLINE=90083302; PubMed=2594783;
RA   Lubahn D.B., Brown T.R., Simental J.A., Higgs H.N., Migeon C.J.,
RA   Wilson E.M., French F.S.;
RT   "Sequence of the intron/exon junctions of the coding region of the
RT   human androgen receptor gene and identification of a point mutation
RT   in a family with complete androgen insensitivity.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:9534-9538(1989).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90258935; PubMed=2342476;
RA   Govindan M.V.;
RT   "Specific region in hormone binding domain is essential for hormone
RT   binding and trans-activation by human androgen receptor.";
RL   Mol. Endocrinol. 4:417-427(1990).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Prostate;
RX   MEDLINE=89017168; PubMed=3174628;
RA   Chang C., Kokontis J., Liao S.;
RT   "Structural analysis of complementary DNA and amino acid sequences of
RT   human and rat androgen receptors.";
RL   Proc. Natl. Acad. Sci. U.S.A. 85:7211-7215(1988).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Prostate;
RX   MEDLINE=89098909; PubMed=2911578;
RA   Tilley W.D., Marcelli M., Wilson J.D., McPhaul M.J.;
RT   "Characterization and expression of a cDNA encoding the human androgen
RT   receptor.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:327-331(1989).
RN   [6]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Prostate;
RX   MEDLINE=91155943; PubMed=2293020;
RA   Marcelli M., Tilley W.D., Wilson C.M., Griffin J.E., Wilson J.D.,
RA   McPhaul M.J.;
RT   "Definition of the human androgen receptor gene structure permits the
RT   identification of mutations that cause androgen resistance: premature
RT   termination of the receptor protein at amino acid residue 588 causes
RT   complete androgen resistance.";
RL   Mol. Endocrinol. 4:1105-1116(1990).
RN   [7]
RP   SEQUENCE OF 189-919 FROM N.A.
RX   MEDLINE=88178111; PubMed=3353726;
RA   Chang C., Kokontis J., Liao S.;
RT   "Molecular cloning of human and rat complementary DNA encoding
RT   androgen receptors.";
RL   Science 240:324-326(1988).
RN   [8]
RP   SEQUENCE OF 468-919 FROM N.A.
RX   MEDLINE=88240407; PubMed=3377788;
RA   Trapman J., Klaassen P., Kuiper G.G.J.M., van der Korput J.A.G.M.,
RA   Faber P.W., van Rooij H.C.J., Geurts van Kessel A., Voorhorst M.M.,
RA   Mulder E., Brinkmann A.O.;
RT   "Cloning, structure and expression of a cDNA encoding the human
RT   androgen receptor.";
RL   Biochem. Biophys. Res. Commun. 153:241-248(1988).
RN   [9]
RP   INTERACTION WITH RAN.
RX   MEDLINE=99329028; PubMed=10400640;
RA   Hsiao P.-W., Lin D.-L., Nakao R., Chang C.;
RT   "The linkage of Kennedy's neuron disease to ARA24, the first
RT   identified androgen receptor polyglutamine region-associated
RT   coactivator.";
RL   J. Biol. Chem. 274:20229-20234(1999).
RN   [10]
RP   POLYMORPHISM OF POLY-GLN REGION.
RX   MEDLINE=92220629; PubMed=1561105;
RA   Sleddens H.F.B.M., Oostra B.A., Brinkmann A.O., Trapman J.;
RT   "Trinucleotide repeat polymorphism in the androgen receptor gene
RT   (AR).";
RL   Nucleic Acids Res. 20:1427-1427(1992).
RN   [11]
RP   POLYMORPHISM OF POLY-GLY REGION.
RC   TISSUE=Blood;
RA   Lu J., Danielsen M.;
RL   Submitted (FEB-1995) to the EMBL/GenBank/DDBJ databases.
RN   [12]
RP   POLYMORPHISM OF POLY-GLN REGION.
RX   MEDLINE=97250535; PubMed=9096391;
RA   Giovannucci E., Stampfer M.J., Krithivas K., Brown M., Brufsky A.,
RA   Talcott J., Hennekens C.H., Kantoff P.W.;
RT   "The CAG repeat within the androgen receptor gene and its
RT   relationship to prostate cancer.";
RL   Proc. Natl. Acad. Sci. U.S.A. 94:3320-3323(1997).
RN   [13]
RP   ERRATUM.
RA   Giovannucci E., Stampfer M.J., Krithivas K., Brown M., Dahl D.,
RA   Brufsky A., Talcott J., Hennekens C.H., Kantoff P.W.;
RL   Proc. Natl. Acad. Sci. U.S.A. 94:8272-8272(1997).
RN   [14]
RP   REVIEW ON VARIANTS.
RX   MEDLINE=93092459; PubMed=1458719;
RA   Pinsky L., Trifiro M.A., Kaufman M., Beitel L.K., Mhatre A.,
RA   Kazemi-Esfarjani P., Sabbaghian N., Lumbroso R., Alvarado C.,
RA   Vasiliou M., Gottlieb B.;
RT   "Androgen resistance due to mutation of the androgen receptor.";
RL   Clin. Invest. Med. 15:456-472(1992).
RN   [15]
RP   REVIEW ON VARIANTS AIS.
RX   MEDLINE=93339360; PubMed=8339746;
RA   Brown T.R., Scherer P.A., Chang Y.-T., Migeon C.J., Ghirri P.,
RA   Murono K., Zhou Z.;
RT   "Molecular genetics of human androgen insensitivity.";
RL   Eur. J. Pediatr. 152 Suppl. 2:S62-S69(1993).
RN   [16]
RP   REVIEW ON VARIANTS.
RX   MEDLINE=94059770; PubMed=8240973;
RA   Sultan C., Lumbroso S., Poujol N., Belon C., Boudon C.,
RA   Lobaccaro J.-M.;
RT   "Mutations of androgen receptor gene in androgen insensitivity
RT   syndromes.";
RL   J. Steroid Biochem. Mol. Biol. 46:519-530(1993).
RN   [17]
RP   REVIEW ON VARIANTS.
RX   MEDLINE=95023089; PubMed=7937057;
RA   Patterson M.N., Hughes I.A., Gottlieb B., Pinsky L.;
RT   "The androgen receptor gene mutations database.";
RL   Nucleic Acids Res. 22:3560-3562(1994).
RN   [18]
RP   REVIEW ON VARIANTS.
RX   MEDLINE=95352489; PubMed=7626493;
RA   Brinkmann A.O., Jenster G., Ris-Stalpers C., van der Korput J.A.G.M.,
```

```
RA   Bruggenwirth H.T., Boehmer A.L.M., Trapman J.;
RT   "Androgen receptor mutations.";
RL   J. Steroid Biochem. Mol. Biol. 53:443-448(1995).
RN   [19]
RP   REVIEW ON VARIANTS.
RX   MEDLINE=97169385; PubMed=9016528;
RA   Gottlieb B., Trifiro M.A., Lumbroso R., Vasiliou D.M., Pinsky L.;
RT   "The androgen receptor gene mutations database.";
RL   Nucleic Acids Res. 25:158-162(1997).
RN   [20]
RP   VARIANT LNCAP ALA-877.
RX   MEDLINE=91083633; PubMed=2260966;
RA   Veldscholte J., Ris-Stalpers C., Kuiper G.G.J.M., Jenster G.,
RA   Berrevoets C.A., Claassen E., van Rooij H.C.J., Trapman J.,
RA   Brinkmann A.O., Mulder E.;
RT   "A mutation in the ligand binding domain of the androgen receptor of
RT   human LNCaP cells affects steroid binding characteristics and
RT   response to anti-androgens.";
RL   Biochem. Biophys. Res. Commun. 173:534-540(1990).
RN   [21]
RP   VARIANTS CAIS CYS-774; GLN-831 AND MET-866.
RX   MEDLINE=91186983; PubMed=2082179;
RA   Brown T.R., Lubahn D.B., Wilson E.M., French F.S., Migeon C.J.,
RA   Corfen J.L.;
RT   "Functional characterization of naturally occurring mutant androgen
RT   receptors from subjects with complete androgen insensitivity.";
RL   Mol. Endocrinol. 4:1759-1772(1990).
RN   [22]
RP   VARIANT CYS-774.
RX   MEDLINE=91310758; PubMed=1856263;
RA   Marcelli M., Tilley W.D., Zoppi S., Griffin J.E., Wilson J.D.,
RA   McPhaul M.J.;
RT   "Androgen resistance associated with a mutation of the androgen
RT   receptor at amino acid 772 (Arg-->Cys) results from a combination of
RT   decreased messenger ribonucleic acid levels and impairment of
RT   receptor function.";
RL   J. Clin. Endocrinol. Metab. 73:318-325(1991).
RN   [23]
RP   VARIANT CAIS PRO-617.
RX   MEDLINE=91154385; PubMed=1999491;
RA   Marcelli M., Zoppi S., Grino P.B., Griffin J.E., Wilson J.D.,
RA   McPhaul M.J.;
RT   "A mutation in the DNA-binding domain of the androgen receptor gene
RT   causes complete testicular feminization in a patient with
RT   receptor-positive androgen resistance.";
RL   J. Clin. Invest. 87:1123-1126(1991).
RN   [24]
RP   VARIANT PAIS CYS-763.
RX   MEDLINE=91185626; PubMed=2010552;
RA   McPhaul M.J., Marcelli M., Tilley W.D., Griffin J.E.,
RA   Isidro-Gutierrez R.F., Wilson J.D.;
RT   "Molecular basis of androgen resistance in a family with a qualitative
RT   abnormality of the androgen receptor and responsive to high-dose
RT   androgen therapy.";
RL   J. Clin. Invest. 87:1413-1421(1991).
RN   [25]

  Query Match             8.5%;  Score 118.5;  DB 1;  Length 919;
  Best Local Similarity   23.1%;  Pred. No. 1.6;
  Matches   69;  Conservative   36;  Mismatches  129;  Indels   65;  Gaps   11;

Qy       1 MELNRSEADEAKAETTPTGGATSSATASGSSSG--RRPRGRPAGSKNKPKPPTIITRDSP 58
           |:| : :  || :|              |||||  |   | |  ||:     |    |:
Db     189 MQLLQQQQQEAVSE-------------GSSSGRAREASGAPTSSKDNYLGGTSTISDNA 234

Qy      59 NVLRSHV-LEVTSGSDISEAVSTYATRRG-CGVCIISGTGAVTNVTIRQPAAPAGGGVIT 116
             |   | : :  | :  | :|     || |    : |      |   | |   | ::
Db     235 KELCKAVSVSMGLGVEALEHLSPGEQLRGDCMYAPLLGVPPAVRPTPCAPLAECKGSLL- 293

Qy     117 LHGRFDILSLTGTALPPPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVL------ 170
             |  :   |       |  ||   |  | :|: :| : : :  :|| : |
Db     294 -----DDSAGKSTEDTAEYSPFKGG---YTKGLEGESLGCSGSAAAGSSGTLELPSTLSL 345

Qy     171 -------MAASFANAVYDRLPIEEEETPPPRT-----TGVQQQQPEASQSSEVTGSGAQA 218
                   ||:: :  |   |:     |||          :: : |     |    : ||
Db     346 YKSGALDEAAAYQSRDYYNFPLALAGPPPPPPPPHPHARIKLENP-LDYGSAWAAAAAQC 404

Qy     219 CESNLQGGNGGGGV------------AFYNLGMNMNNFQFSGGDIYGMSGGSGGGGGG 264
             :|   :| |                ::: |      |    | :||  || ||||||
Db     405 RYGDLASLHGAGAAGPGSGSPSAAASSSWHTL-----FTAEEGQLYGPCGGGGGGGGG 457


RESULT  11
BET3_MESAU
ID   BET3_MESAU     STANDARD;      PRT;   367 AA.
AC   O09029;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   BETA3 protein.
OS   Mesocricetus auratus (Golden hamster).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC   Mesocricetus.
OX   NCBI_TaxID=10036;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96140430; PubMed=8552091;
RA   Peyton M., Stellrecht C.M.M., Naya F.J., Huang H.-P., Samora P.J.,
RA   Tsai M.-J.;
RT   "BETA3, a novel helix-loop-helix protein, can act as a negative
RT   regulator of BETA2 and MyoD-responsive genes.";
RL   Mol. Cell. Biol. 16:626-633(1996).
CC   -!- FUNCTION: INHIBITS DNA BINDING OF TCF3 (E47) HOMODIMERS AND TCF3
CC       (E47) / NEUROD1 HETERODIMERS AND ACTS AS A STRONG REPRESSOR OF
CC       NEUROD1 AND MYOD-RESPONSIVE GENES, PROBABLY BY HETERODIMERIZATION
CC       WITH CLASS A BASIC HELIX-LOOP-HELIX FACTORS. DESPITE THE PRESENCE
CC       OF AN INTACT BASIC DOMAIN, DOES NOT BIND TO DNA.
CC   -!- SUBUNIT: HETERODIMER WITH OTHER BHLH PROTEINS, LIKE TCF3 (E47).
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- TISSUE SPECIFICITY: KIDNEY, LUNG, BRAIN AND PANCREAS (INSULINOMA).
CC   -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC       TRANSCRIPTION FACTORS. "ATONAL" SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; S80870; AAB50691.1; -.
DR   InterPro; IPR003015; HLH_Myc.
DR   InterPro; IPR001092; HLH_dim.
DR   Pfam; PF00010; HLH; 1.
DR   SMART; SM00353; HLH; 1.
DR   PROSITE; PS00038; HELIX_LOOP_HELIX; 1.
KW   Nuclear protein; Transcription regulation; Repressor.
FT   DOMAIN       11     14       POLY-ALA.
FT   DOMAIN       58     62       POLY-SER.
FT   DOMAIN       83     99       POLY-GLY.
FT   DOMAIN      174    179       POLY-GLY.
FT   DOMAIN      204    217       POLY-GLY.
FT   DNA_BIND    229    240       BASIC DOMAIN.
FT   DOMAIN      241    282       HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT   DOMAIN      311    319       POLY-ALA.
SQ   SEQUENCE   367 AA;  35905 MW;  6CAB9AFF96E85F77 CRC64;


  Query Match             8.4%;  Score 117;  DB 1;  Length 367;
  Best Local Similarity   23.3%;  Pred. No. 0.77;
  Matches   68;  Conservative   17;  Mismatches   65;  Indels  142;  Gaps   15;

Qy      10 EAKAETTPTGGATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDSPNVLRSHVLEVT 69
           ||   :|| |   | |          || ||| | : |
```

```
Db      34 EAAFRSTPPGMDLSLAPP--------PRERPASSSSSP---------------------- 63

Qy      70 SGSDISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGRFDILSLTGT 129
                      ||                  :|| | | |::
Db      64 --------------LGC---------------FEPADPEGAGLL-------------- 78

Qy     130 ALPPPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVYDRLP----- 184
           ||||   | ||     ||| |   ||   | :   | :|  |    :     ||
Db      79 -LPPPG--GGGG-----AGGGG---GGGGGGGVSVPGLLVGSAGVGGDPNLSSLPAGAAL 127

Qy     185 ------------IEE----EETPPPRTTG-------VQQQQPEASQSSEVTGSGAQACE- 220
                     : |    |::|   : |              | ||  :   | | :  |
Db     128 CLKYGESAGRGSVAESSGGEQSPDDDSDGRCELVLRAGGADPRASPGA--GGGGTKVVEG 185

Qy     221 ---SNLQGGNG---GGGVAFYNLGMNMNNFQFSGGDIYGMSGGSGGGGGGGAT 266
               ::| || |   ||                |||      ||||||||| :
Db     186 CSNAHLHGGAGLPPGGSTG-------------SGG------GGSGGGGGGGS 218


RESULT  12
CIKE_DROME
ID   CIKE_DROME     STANDARD;      PRT;  1174 AA.
AC   Q02280;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Potassium channel protein eag.
GN   EAG.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91262635; PubMed=1840699;
RA   Warmke J., Drysdale R., Ganetzky B.;
RT   "A distinct potassium channel polypeptide encoded by the Drosophila
RT   eag locus.";
RL   Science 252:1560-1562(1991).
CC   -!- FUNCTION: PROTEIN EAG IS MOST PROBABLY A STRUCTURAL COMPONENT OF
CC       THE POTASSIUM CHANNEL AND MEDIATES THE POTASSIUM PERMEABILITY OF
CC       MEMBRANES.
CC   -!- SUBUNIT: DIMER (POTENTIAL).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- MISCELLANEOUS: THE SEGMENT S4 IS PROBABLY THE VOLTAGE-SENSOR AND
CC       IS CHARACTERIZED BY A SERIES OF POSITIVELY CHARGED AMINO ACIDS AT
CC       EVERY THIRD POSITION.
CC   -!- MISCELLANEOUS: THE SEGMENT H5 IS THOUGHT TO LINE THE CHANNEL PORE.
CC   -!- SIMILARITY: TO THE MEMBERS OF THE POTASSIUM CHANNEL PROTEINS
CC       OF THE SH SUPERFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 CYCLIC NUCLEOTIDE-BINDING DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M61157; AAA28495.1; -.
DR   PIR; A40853; A40853.
DR   FlyBase; FBgn0000535; eag.
DR   InterPro; IPR000636; Cation_chan_non_lig.
DR   InterPro; IPR001622; Channel_pore_K.
DR   InterPro; IPR001610; PAC.
DR   InterPro; IPR000700; PAS-assoc_C.
DR   InterPro; IPR000014; PAS.
DR   InterPro; IPR000595; cNMP_binding.
DR   Pfam; PF00027; cNMP_binding; 1.
DR   Pfam; PF00520; ion_trans; 1.
DR   Pfam; PF00785; PAC; 1.
DR   SMART; SM00086; PAC; 1.
DR   SMART; SM00091; PAS; 1.
DR   SMART; SM00100; cNMP; 1.
DR   PROSITE; PS50042; CNMP_BINDING_3; 1.
DR   PROSITE; PS50112; PAS; 1.
DR   PROSITE; PS50113; PAC; 1.
KW   Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
KW   Glycoprotein; Phosphorylation; Repeat.
FT   TRANSMEM    227    246       SEGMENT S1 (POTENTIAL).
FT   TRANSMEM    269    291       SEGMENT S2 (POTENTIAL).
FT   TRANSMEM    314    335       SEGMENT S3 (POTENTIAL).
FT   TRANSMEM    343    369       SEGMENT S4 (POTENTIAL).
FT   TRANSMEM    372    393       SEGMENT S5 (POTENTIAL).
FT   TRANSMEM    442    467       SEGMENT H5 (POTENTIAL).
FT   TRANSMEM    471    493       SEGMENT S6 (POTENTIAL).
FT   DOMAIN       43     97       PAS.
FT   DOMAIN      113    165       PAC.
FT   NP_BIND     571    688       CNMP.
FT   CARBOHYD    262    262       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    412    412       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    424    424       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1174 AA;  126236 MW;  344C80DC06E4340E CRC64;


  Query Match             8.4%;  Score 117;  DB 1;  Length 1174;
  Best Local Similarity   21.6%;  Pred. No. 2.5;
  Matches   58;  Conservative   23;  Mismatches   94;  Indels   94;  Gaps    8;

Qy      20 GATSSATASGSSSGRRPRGRP---------AGSKNKPKPPTIITRDSPNVLRSHVLEVTS 70
           ||:|   |  :|||:   |             |:     |||
Db     932 GASSGGNAPDNSSGQTTPGDEICAGCGAGGGGTPTTQAPPT------------------ 972

Qy      71 GSDISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGRFDILSLTGTA 130
             | ::  | |  |    |    ||:|: |       | |||  :   |   : | |
Db     973 -SAVTSPVDTVITISSPGA---SGSGSGTGAGAGSAVAGAGGAGLLDPGATVVSSAGGNG 1028

Qy     131 LPP------------PAPPGAGGLTVYLAGGQGQVVGGNVAGS-LIASGPVVLMAASFA 176
           | |             ||||     |:    |        |||| : :| |
Db    1029 LGPLMLKKRRSKSGKAPAPPEQ---TLASTAGTATAAPAGVAGSGMTSSAP--------- 1076

Qy     177 NAVYDRLPIEEEETPPPRTTGVQQQQPEASQSSEVTGSGAQACESNLQGGNGGGGVAFYN 236
                          :  |||| :::     |  ||:
Db    1077 -----------------ASADQQQQHQSAADQSPTTPGAEL----------------- 1100

Qy     237 LGMNMNNFQFSGGDIYGMSGGSGGGGGGA 265
           | : :   |:   :   | |  |||||:
Db    1101 LHLRLLEEDFTAAQLPSTSSGGAGGGGGS 1129


RESULT  13
IRS1_MOUSE
ID   IRS1_MOUSE     STANDARD;      PRT;  1233 AA.
AC   P35569;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Insulin receptor substrate-1.
GN   IRS1 OR IRS-1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94220494; PubMed=8167159;
RA   Araki E., Haag B.L. III, Kahn C.R.;
RT   "Cloning of the mouse insulin receptor substrate-1 (IRS-1) gene and
RT   complete sequence of mouse IRS-1.";
RL   Biochim. Biophys. Acta 1221:353-356(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
```

```
RX   MEDLINE=93192326; PubMed=8448209;
RA   Keller S.R., Aebersold R., Garner C.W., Lienhard G.E.;
RT   "The insulin-elicited 160 kDa phosphotyrosine protein in mouse
RT   adipocytes is an insulin receptor substrate 1: identification by
RT   cloning.";
RL   Biochim. Biophys. Acta 1172:323-326(1993).
CC   -!- FUNCTION: MAY MEDIATE THE CONTROL OF VARIOUS CELLULAR PROCESSES
CC       BY INSULIN. WHEN PHOSPHORYLATED BY THE INSULIN RECEPTOR BINDS
CC       SPECIFICALLY TO VARIOUS CELLULAR PROTEINS CONTAINING SH2 DOMAINS
CC       SUCH AS PHOSPHATIDYLINOSITOL 3-KINASE P85 SUBUNIT OR GRB-2.
CC   -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 PTB DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L24563; AAA39335.1; -.
DR   EMBL; X69722; CAA49378.1; -.
DR   PIR; S43514; S43514.
DR   HSSP; P35568; 1IRS.
DR   MGD; MGI:99454; Irs1.
DR   InterPro; IPR002404; Insulin_Recep_S-1.
DR   InterPro; IPR001849; PH.
DR   Pfam; PF02174; IRS; 1.
DR   Pfam; PF00169; PH; 1.
DR   PRINTS; PR00628; INSULINRSI.
DR   SMART; SM00233; PH; 1.
DR   SMART; SM00310; PTBI; 1.
DR   PROSITE; PS50003; PH_DOMAIN; 1.
KW   Phosphorylation.
FT   DOMAIN       12    115       PH.
FT   DOMAIN      152    262       PTB.
FT   DOMAIN      675    680       POLY-SER.
FT   DOMAIN      872    877       POLY-GLN.
FT   DOMAIN     1119   1128       POLY-GLY.
FT   DOMAIN     1194   1198       POLY-PRO.
FT   MOD_RES      99     99       PHOSPHORYLATION (BY CK2)
FT                                (BY SIMILARITY).
FT   MOD_RES     460    460       PHOSPHORYLATION (BY INSR)
FT                                (BY SIMILARITY).
FT   MOD_RES     608    608       PHOSPHORYLATION (BY INSR)
FT                                (BY SIMILARITY).
FT   MOD_RES     628    628       PHOSPHORYLATION (BY INSR)
FT                                (BY SIMILARITY).
FT   MOD_RES     891    891       PHOSPHORYLATION (BY INSR)
FT                                (BY SIMILARITY).
FT   MOD_RES     935    935       PHOSPHORYLATION (BY INSR)
FT                                (BY SIMILARITY).
FT   MOD_RES     983    983       PHOSPHORYLATION (BY INSR)
FT                                (BY SIMILARITY).
FT   MOD_RES    1173   1173       PHOSPHORYLATION (BY INSR)
FT                                (BY SIMILARITY).
FT   MOD_RES    1220   1220       PHOSPHORYLATION (BY INSR)
FT                                (BY SIMILARITY).
FT   CONFLICT   1038   1039       MISSING (IN REF. 2).
FT   CONFLICT   1182   1182       H -> R (IN REF. 2).
SQ   SEQUENCE   1233 AA;  130723 MW;  C0E9B2D890DADD87 CRC64;


  Query Match             8.4%;  Score 116.5;  DB 1;  Length 1233;
  Best Local Similarity   22.8%;  Pred. No. 2.8;
  Matches   84;  Conservative   29;  Mismatches  125;  Indels  131;  Gaps   16;

Qy       5 RSEADEAKAETTPTGGATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDSPNVLRSH 64
           |  :   :   | |   :||:|:| |  |     ||  |   |               |
Db     785 RLSSSSGRLRYTATAEDSSSSTSSDSLGGGYCGARPESSLTHP-------------HHH 830

Qy      65 VLEVTSGSDISEAVSTYA-----TRRGCGVCIISGTGAVTNVTIRQ----------PAAP 109
           ||:      :  |  | :     ||   |        | |   :|:          |  |
Db     831 VLQPHLPRKVDTAAQTNSRLARPTRLSLG-----DPKASTLPRVREQQQQQQSSLHPPEP 885

Qy     110 AGGGVIT----------LHGRFDILSLTGTALPP------------------------ 133
             |                | |      |      ||
Db     886 KSPGEYVNIEFGSGQPGYLAGPATSRSSPSVRCPPQLHPAPREETGSEEYMNMDLGPGRR 945

Qy     134 --------------PAPPGAGGL---TVYLAGGQGQVVG---GNVAGSLIASGPVVLMA 172
                         |||||:  :   |  :   :|  :    |    | : : ||
Db     946 ATWQESGGVELGRIGPAPPGSATVCRPTRSVPNSRGDYMTMQIGCPRQSYVDTSPVA--P 1003

Qy     173 ASFANAVYDRLPIEEEETPPPRTTG-VQQQQPEASQSSEVTGSGA---QACESNLQGG-N 227
           |:|:    |  |  |:   || ||        || |: ||  ||   ||  |:| ||
Db    1004 VSYADM---RTGIAAEKASLPRPTGAAPPPSSTASSSASVTPQGATAEQATHSSLLGGPQ 1060

Qy     228 GGGGV-AFYNLGMNMNNFQ--------------------FSGGDIYGMS---------GG 257
           | ||: ||  : :: |: |                    ||     | :          ||
Db    1061 GPGGMSAFTRVNLSPNHNQSAKVIRADTQGCRRRHSSETFSAPTRAGNTVPFGAGAAVGG 1120

Qy     258 SGGGGGGAT 266
           |||||||  :
Db    1121 SGGGGGGGS 1129


RESULT  14
SUS_DROME
ID   SUS_DROME      STANDARD;      PRT;  1322 AA.
AC   P22293;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Suppressor of sable protein.
GN   SU(S).
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OREGON-R;
RX   MEDLINE=91117256; PubMed=1703632;
RA   Voelker R.A., Gibson W., Graves J.P., Sterling J.F., Eisenberg M.T.;
RT   "The Drosophila suppressor of sable gene encodes a polypeptide with
RT   regions similar to those of RNA-binding proteins.";
RL   Mol. Cell. Biol. 11:894-905(1991).
RN   [2]
RP   SEQUENCE OF 1-9 FROM N.A.
RX   MEDLINE=91169252; PubMed=1963868;
RA   Voelker R.A., Graves J.P., Gibson W., Eisenberg M.T.;
RT   "Mobile element insertions causing mutations in the Drosophila
RT   suppressor of sable locus occur in DNase I hypersensitive subregions
RT   of 5'-transcribed nontranslated sequences.";
RL   Genetics 126:1071-1082(1990).
CC   -!- FUNCTION: AFFECTS THE TRANSCRIPT LEVELS OF THOSE ALLELES THAT IT
CC       SUPPRESSES. MAY BE INVOLVED IN RNA METABOLISM.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- DEVELOPMENTAL STAGE: AT ALL STAGES.
CC   -!- SIMILARITY: HAS REGIONS SIMILAR TO THOSE OF RNA-BINDING PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M57889; AAA28920.1; -.
DR   EMBL; X59364; CAA42010.1; -.
DR   PIR; A39612; A39612.
DR   FlyBase; FBgn0003575; su(s).
DR   InterPro; IPR000571; Zf-CCCH.
```

```
DR   Pfam; PF00642; zf-CCCH; 2.
KW   RNA-binding; Nuclear protein.
FT   DOMAIN      138    327       HIGHLY CHARGED DOMAIN.
FT   DOMAIN      446    474       GLN-RICH (OPA-REPEAT).
FT   DOMAIN     1087   1162       RNA-BINDING (BY SIMILARITY).
SQ   SEQUENCE   1322 AA;  143555 MW;  D5F534EB5702EA08 CRC64;


  Query Match             8.3%;  Score 116;  DB 1;  Length 1322;
  Best Local Similarity   22.1%;  Pred. No. 3.2;
  Matches   71;  Conservative   26;  Mismatches  108;  Indels  116;  Gaps   12;

Qy      17 PTGGATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDSPNVLRSHVLEVTSGSDISE 76
           || ||::: | : ::||        |||                  : |: |   :|
Db    1036 PTKGASTANTTAPNASG--------GSK------------------QISEIRSLLQVSN 1068

Qy      77 AVSTYATRRGCGV----------------------CIISGTGAVTNVTIRQPAAPAGGGV 114
             :   |     |                      |   :  | | |::|  |
Db    1069 WYNNLGTNNKIMVNQQLALVFTELKKFHQLPNDAPKIFDVSFIVNNTTLQQIFAKL---- 1124

Qy     115 ITLHGRFDILSLTGTALPPPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAAS 174
           | :   |  :  |  |   |  :   |  |  |||   |     | |||   |
Db    1125 ------FIFVDDNGEVVQIPEEPNGNGAALGGGGDSGGGVGGGGGG-----GGVVLPNLS 1173

Qy     175 FANAVYDRLPIEEEETPPP-----RTTGVQQQQ----PEASQ--------SSEVTGSGAQ 217
           ::       |||     | :|:  |    |  :|              |:|
Db    1174 QPPPNLSQM----LRLPPPNIRMLRMSGMMMQMGNVGPPFNQPPPRGGLMGMPPNGNGLN 1229

Qy     218 ACESNLQ----------------------GGNGGGGV-------AFYNLGMNMNNFQFS 247
             ||                        ||| |||         :  |:||  |||  :
Db    1230 QGVGNLGGLGQLGINQGGGPVPNGNPFNPFGGNNGGGAGVMNNMNSMGNMGMGFNNFNNN 1289

Qy     248 GG-DIYGMSGGSGGGGGGATR 267
           ||   :   |||||  | |  |
Db    1290 GGRGGHFPGGGSGGNGNGNNR 1310


RESULT  15
CSP_PLACB
ID   CSP_PLACB      STANDARD;      PRT;   378 AA.
AC   P08672;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-JAN-1988 (Rel. 06, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Circumsporozoite protein precursor (CS).
OS   Plasmodium cynomolgi (strain Berok).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5828;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87102878; PubMed=3802196;
RA   Galinski M.R., Arnot D.E., Cochrane A.H., Barnwell J.W.,
RA   Nussenzweig R.S., Enea V.;
RT   "The circumsporozoite gene of the Plasmodium cynomolgi complex.";
RL   Cell 48:311-319(1987).
CC   -!- DOMAIN: THERE ARE 10 TANDEM COPIES OF A 9-RESIDUE REPEAT (PRECEDED
CC       BY A 6-RESIDUE INCOMPLETE REPEAT) AND 3 TANDEM COPIES OF A 16-
CC       RESIDUE REPEAT (FOLLOWED BY 3 SHORTER, INCOMPLETE COPIES).
CC   -!- MISCELLANEOUS: THE CIRCUMSPOROZOITE PROTEIN IS THE IMMUNODOMINANT
CC       SURFACE ANTIGEN ON THE SPOROZOITE (THE INFECTIVE STAGE OF THE
CC       MALARIA PARASITE THAT IS TRANSMITTED FROM THE MOSQUITO TO THE
CC       VERTEBRATE HOST).
CC   -!- MISCELLANEOUS: THE C-TERMINAL REGION IS PROBABLY USED FOR
CC       ANCHORING THE PROTEIN TO THE CELL MEMBRANE. THE REPEAT SEQUENCES
CC       WOULD BE THE SURFACE ANTIGEN OF THE ORGANISM.
CC   -!- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M15104; AAA29532.1; -.
DR   PIR; D26255; OZZQAB.
DR   InterPro; IPR003067; Crcmsprzoite.
DR   InterPro; IPR000884; TSP1.
DR   Pfam; PF00090; tsp_1; 1.
DR   PRINTS; PR01303; CRCMSPRZOITE.
DR   SMART; SM00209; TSP1; 1.
KW   Malaria; Sporozoite; Repeat; Signal.
FT   SIGNAL        1     19       PROBABLE.
FT   CHAIN        20    378       CIRCUMSPOROZOITE PROTEIN.
FT   DOMAIN       97    192       10.5 X 9 AA REPEATS.
FT   DOMAIN      193    240       3 X 16 AA TANDEM REPEATS.
FT   REPEAT      241    251
FT   REPEAT      252    260
FT   REPEAT      261    268
SQ   SEQUENCE   378 AA;  36286 MW;  779BA081C140793F CRC64;


  Query Match             8.3%;  Score 115.5;  DB 1;  Length 378;
  Best Local Similarity   24.2%;  Pred. No. 0.98;
  Matches   64;  Conservative    9;  Mismatches   99;  Indels   93;  Gaps   10;

Qy       7 EADEAKAETTPTG-GATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDSPNVLRSHV 65
           | | | |   | | || ::    |: :     | ||       |       :|
Db     104 EGDGAPA--APAGDGAPAAPAGDGAPAAPAGDGAPAAPAGDGAPAAPAGDGAP------- 154

Qy      66 LEVTSGSDISEAVSTYATRRGCGV-CIISGTGAVTNVTIRQPAAPAGGGVITLHGRFDIL 124
                        |   | |     :| ||        ||||| |
Db     155 ----------------AAPAGDGAPAAPAGDGA--------PAAPAGDGA---------- 180

Qy     125 SLTGTALPPPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVYDRLP 184
                 | || | |       ||||    ||| ||   | |
Db     181 --------PAAPAGDGAPAGNRAGGQ-PAAGGNQAGGNRAGG------------------ 213

Qy     185 IEEEETPPPRTTGVQQQQPEA--SQSSEVTGSGAQACESNLQGGN--GGGGVAFYNLGMN 240
                         || |  :|:     |  |   |  ||    ||  |    | |
Db     214 ----------------QPAAGGNQAGGNRAGGQPAAGGNQAGGQPAAGGNQAGAQAGGN 256

Qy     241 MNNFQFSGGDIYGMSGGSGGGGGGA 265
             |  | :      |  ||  ||
Db     257 QAGAQAGGANAGNKKAGEAGGNAGA 281


Search completed: July 10, 2002, 22:31:08
Job time: 585 sec
```

```
Qy       1 MELNRSEADEAKAETTPTGGATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDSPNV 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MELNRSEADEAKAETTPTGGATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDSPNV 60

Qy      61 LRSHVLEVTSGSDISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGR 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 LRSHVLEVTSGSDISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGR 120

Qy     121 FDILSLTGTALPPPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVY 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 FDILSLTGTALPPPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVY 180

Qy     181 DRLPIEEEETPPPRTTGVQQQQPEASQSSEVTGSGAQACESNLQGGNGGGGVAFYNLGMN 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 DRLPIEEEETPPPRTTGVQQQQPEASQSSEVTGSGAQACESNLQGGNGGGGVAFYNLGMN 240

Qy     241 MNNFQFSGGDIYGMSGGSGGGGGGATRPAF 270
           ||||||||||||||||||||||||||||||
Db     241 MNNFQFSGGDIYGMSGGSGGGGGGATRPAF 270
```

RESULT 2
Q9S7C9
ID Q9S7C9 PRELIMINARY; PRT; 311 AA.
AC Q9S7C9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE ESCAROLA.
GN ESC OR F9H16.12.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Weigel D., Ahn J.H., Blazquez M.A., Borevitz J., Christensen S.K.,
RA Fankhauser C., Ferrandiz C., Kardailsky I., Neff M.M., Nguyen J.T.,
RA Sato S., Wang Z., Xia Y., Dixon R.A., Harrison M.J., Lamb C.J.,
RA Yanofsky M.F., Chory J.;
RT "Activation Tagging in Arabidopsis.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,
RA Altafi H., Araujo R., Huizar L., Rowley D., Buehler E., Dunn P.,
RA Gonzalez A., Kremenetskaia I., Kim C., Lenz C., Li J., Liu S.,
RA Luros S., Schwartz J., Shinn P., Toriumi M., Vysotskaia V.S.,
RA Walker M., Yu G., Ecker J., Theologis A., Davis R.W.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF194974; AAF07197.1; -.
DR EMBL; AC007369; AAD30602.1; -.
DR InterPro; IPR000637; AT_hook.
DR Pfam; PF02178; AT_hook; 1.
DR SMART; SM00384; AT_hook; 1.
SQ SEQUENCE 311 AA; 31842 MW; A80B445C9776EB7D CRC64;

Query Match 50.8%; Score 706.5; DB 10; Length 311;
Best Local Similarity 53.0%; Pred. No. 5e-41;
Matches 151; Conservative 25; Mismatches 56; Indels 53; Gaps 6;

```
Qy       7 EADEAKAETTPTGGATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDSPNVLRSHVL 66
           |:| :  :   |  |    | |: |:|||||| ||||| ||| |:|||||| ||||||
Db      59 ESDHSNKDHHQQGRPDSDPNTSSSAPGKRPRGRPPGSKNKAKPPIIVTRDSPNALRSHVL 118

Qy      67 EVTSGSDISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAP-------AGGGVITLHG 119
           ||: |:|| |:||||| ||| || :: | | |:|||:|||  |        ||||:||||
Db     119 EVSPGADIVESVSTYARRRGRGVSVLGGNGTVSNVTLRQPVTPGNGGGVSGGGGVVTLHG 178

Qy     120 RFDILSLTGTALPPPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAV 179
```

```
           ||:||||||| |||||||||||:::|||||||||||:|   |||| ||:||||||:|||
Db     179 RFEILSLTGTVLPPPAPPGAGGLSIFLAGGQGQVVGGSVVAPLIASAPVILMAASFSNAV 238

Qy     180 YDRLPIEEEETP----------PPRTTGVQQQQPEASQSSEVTGSGAQACESNLQGGNG 228
           ::||||||||            ||:     || | ||  | ||| |         |||
Db     239 FERLPIEEEEEEGGGGGGGGGGPPQ----MQQAPSASPPSGVTGQGQ-------LGGNV 287

Qy     229 GGGVAFYNLGMNMNNFQFSGGDIYGMSGGS---GGGGGGATRPAF 270
           ||                 || ||     | | |  :|| |
Db     288 GG-------------------YGFSGDPHLLGWGAGTPSRPPF 311
```

RESULT 3
O23620
ID O23620 PRELIMINARY; PRT; 292 AA.
AC O23620;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 29.7 KDA PROTEIN.
GN AT4G17800.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98121113; PubMed=9461215;
RA Bevan M., Bancroft I., Bent E., Love K., Goodman H., Dean C.,
RA Bergkamp R., Dirkse W., van Staveren M., Stiekema W., Drost L.,
RA Ridley P., Hudson S.A., Patel K., Murphy G., Piffanelli P., Wedler H.,
RA Wedler E., Wambutt R., Weitzenegger T., Pohl T.M., Terryn N.,
RA Gielen J., Villarroel R., De Clerck R., van Montagu M., Lecharny A.,
RA Auborg S., Gy I., Kreis M., Lao N., Kavanagh T., Hempel S., Kotter P.,
RA Entian K.D., Rieger M., Schaeffer M., Funk B., Mueller-Auer S.,
RA Silvey M., James R., Montfort A., Pons A., Puigdomenech P., Douka A.,
RA Voukelatou E., Milioni D., Hatzopoulos P., Piravandi E., Obermaier B.,
RA Hilbert H., Duesterhoft A., Moores T., Jones J.D.G., Eneva T.,
RA Palme K., Benes V., Rechman S., Ansorge W., Cooke R., Berger C.,
RA Delseny M., Voet M., Volckaert G., Mewes H.W., Klosterman S.,
RA Schueller C., Chalwatzis N.;
RT "Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of
RT Arabidopsis thaliana.";
RL Nature 391:485-488(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; Z97344; CAB10560.1; -.
DR EMBL; AL161547; CAB78783.1; -.
KW Hypothetical protein.
SQ SEQUENCE 292 AA; 29724 MW; 2A948700FA543C4F CRC64;

Query Match 46.4%; Score 646.5; DB 10; Length 292;
Best Local Similarity 52.7%; Pred. No. 5.9e-37;
Matches 136; Conservative 32; Mismatches 47; Indels 43; Gaps 8;

```
Qy      19 GGATSSATASGSSS-----GRRPRGRPAGSKNKPKPPTIITRDSPNVLRSHVLEVTSGSD 73
           ||: ||    |         |||||||| |||||||| ||||:| | ||:|:||||:| |
Db      62 GGSGSSGGGGGHGGGGDVVGRRPRGRPPGSKNKPKPPVIITRESANTLRAHILEVTNGCD 121

Qy      74 ISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGRFDILSLTGTALPP 133
           : : |:||| ||  |:|::||:| ||||:||||:  | | |:|| | |:||||:|: |||
Db     122 VFDCVATYARRRQRGICVLSGSGTVTNVSIRQPS--AAGAVVTLQGTFEILSLSGSFLPP 179

Qy     134 PAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVYDRLPIEEEETPPP 193
           ||||||  ||::|||||||||||:| | | |:|||:::|||| |  |:|||:||:|
Db     180 PAPPGATSLTIFLAGGQGQVVGGSVVGELTAAGPVIVIAASFTNVAYERLPLEEDE---- 235

Qy     194 RTTGVQQQQPEASQSSEVTGSGAQACESNLQGGN--------GGGGVAFYNLGMNMN-NF 244
                ||||          | |:        |||        ||||: |:|| |||  |
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 10, 2002, 22:19:08 ; Search time 103.68 Seconds
(without alignments)
450.508 Million cell updates/sec

Title: US-09-823-676-2
Perfect score: 1392
Sequence: 1 MELNRSEADEAKAETTPTGG.....IYGMSGGSGGGGGGATRPAF 270

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1392	100.0	270	10	O65489	O65489 arabidopsis
2	706.5	50.8	311	10	Q9S7C9	Q9s7c9 arabidopsis
3	646.5	46.4	292	10	O23620	O23620 arabidopsis
4	645.5	46.4	302	10	Q9C9K7	Q9c9k7 arabidopsis
5	637.5	45.8	315	10	Q9SR17	Q9sr17 arabidopsis
6	629.5	45.2	317	10	O22130	O22130 arabidopsis
7	625.5	44.9	339	10	Q9SZ70	Q9sz70 arabidopsis
8	616	44.3	285	10	O82166	O82166 arabidopsis
9	605	43.5	324	10	O49662	O49662 arabidopsis
10	581	41.7	265	10	Q9LZX7	Q9lzx7 arabidopsis
11	572	41.1	310	10	Q9M2S3	Q9m2s3 arabidopsis
12	554	39.8	257	10	Q9SJG4	Q9sjg4 arabidopsis
13	499.5	35.9	276	10	Q9LTA2	Q9lta2 arabidopsis
14	417	30.0	206	10	Q9M9R4	Q9m9r4 arabidopsis
15	407	29.2	265	10	Q941Y9	Q941y9 oryza sativ
16	289.5	20.8	404	10	Q9LVB0	Q9lvb0 arabidopsis
17	289	20.8	771	10	Q9XHY4	Q9xhy4 oryza sativ
18	283	20.3	405	10	Q9AYM0	Q9aym0 oryza sativ
19	282	20.3	351	10	Q94F52	Q94f52 arabidopsis
20	274	19.7	296	10	O22812	O22812 arabidopsis
21	274	19.7	351	10	O23142	O23142 arabidopsis
22	264.5	19.0	419	10	Q9FHM5	Q9fhm5 arabidopsis
23	259	18.6	826	10	Q9SH33	Q9sh33 arabidopsis
24	258.5	18.6	300	10	Q9ZRR7	Q9zrr7 antirrhinum
25	255.5	18.4	334	10	O49658	O49658 arabidopsis
26	248.5	17.9	386	10	Q9FIR1	Q9fir1 arabidopsis
27	246.5	17.7	347	10	O04696	O04696 pisum sativ
28	243	17.5	439	10	O49350	O49350 arabidopsis
29	243	17.5	439	10	Q940I0	Q940i0 arabidopsis
30	242.5	17.4	334	10	O04695	O04695 pisum sativ
31	239.5	17.2	404	10	Q9SB31	Q9sb31 arabidopsis
32	239	17.2	455	10	O49694	O49694 arabidopsis
33	227	16.3	365	10	Q9SZ73	Q9sz73 arabidopsis
34	225	16.2	348	10	O80834	O80834 arabidopsis
35	221	15.9	418	10	Q9SR16	Q9sr16 arabidopsis
36	218	15.7	345	10	O81321	O81321 arabidopsis
37	211.5	15.2	348	10	Q9M2D3	Q9m2d3 arabidopsis
38	162.5	11.7	574	10	Q9SJQ2	Q9sjq2 arabidopsis
39	148	10.6	848	5	Q9V889	Q9v889 drosophila
40	146	10.5	405	5	Q23057	Q23057 caenorhabdi
41	143.5	10.3	818	6	Q9N1P0	Q9n1p0 bos taurus
42	139	10.0	1468	5	Q9GUB5	Q9gub5 galleria me
43	138.5	9.9	265	5	Q23347	Q23347 caenorhabdi
44	137	9.8	694	16	O53212	O53212 mycobacteri
45	136.5	9.8	1610	16	Q92KQ8	Q92kq8 rhizobium m

ALIGNMENTS

RESULT 1
O65489
ID O65489 PRELIMINARY; PRT; 270 AA.
AC O65489;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 26.8 KDA PROTEIN.
GN F23E12.50 OR AT4G35390.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA Hoheisel J., Jesse T., Heijnen L., Vos P., Mewes H.W., Mayer K.F.X.,
RA Schueller C.;
RL Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AL022604; CAA18730.1; -.
DR EMBL; AL161587; CAB80256.1; -.
KW Hypothetical protein.
SQ SEQUENCE 270 AA; 26774 MW; 4C21487C2719E114 CRC64;

Query Match 100.0%; Score 1392; DB 10; Length 270;
Best Local Similarity 100.0%; Pred. No. 4.7e-88;
Matches 270; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Db    236 -----QQQQ---------LGGGSNG------GGNLFPEVAAGGGGGLPFFNLPMNMQPNV 275

Qy    245 QFSGGDIYGMSGGSGGGG 262
          |     : |  | ||| |
Db    276 QL---PVEGWPGNSGGRG 290


RESULT   4
Q9C9K7
ID   Q9C9K7      PRELIMINARY;      PRT;   302 AA.
AC   Q9C9K7;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   HYPOTHETICAL 30.6 KDA PROTEIN.
GN   F14G6.10.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RX   MEDLINE=21016719; PubMed=11130712;
RA   Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA   White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA   Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA   Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA   Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA   Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA   Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA   Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA   Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA   Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziali A.,
RA   Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA   Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA   Sakano H., Salzerg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA   Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA   Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA   Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT   "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 408:816-820(2000).
DR   EMBL; AC015450; AAG51949.1; -.
DR   InterPro; IPR000637; AT_hook.
DR   Pfam; PF02178; AT_hook; 1.
DR   SMART; SM00384; AT_hook; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   302 AA;  30615 MW;  6EC755021E3DD375 CRC64;


  Query Match             46.4%;  Score 645.5;  DB 10;  Length 302;
  Best Local Similarity   49.5%;  Pred. No. 7.2e-37;
  Matches  138;  Conservative   33;  Mismatches   63;  Indels   45;  Gaps    7;

Qy      9 DEAKAETTPTGGATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDSPNVLRSHVLEV 68
          ||: :   |      |   :  |:|:|||||| ||||||||| |:|||||||||||||||
Db     52 DESDSNKDP-----GSDPVTSGSTGKRPRGRPPGSKNKPKPPVIVTRDSPNVLRSHVLEV 106

Qy     69 TSGSDISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAG-------GGVITLHGRF 121
          :||:|| |:|:||| ||| || |:|| | | ||::|||| |        |||: |||||
Db    107 SSGADIVESVTTYARRRGRGVSILSGNGTVANVSLRQPATTAAHGANGGTGGVVALHGRF 166

Qy    122 DILSLTGTALPPPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVYD 181
          :|||||| |||||||:|||:::|:| ||||:||||   |:|||||:||||||:|| ::
Db    167 EILSLTGTVLPPPAPPGSGGLSIFLSGVQGQVIGGNVVAPLVASGPVILMAASFSNATFE 226

Qy    182 RLPIEEEETPPPRTTGVQQQQPEASQSSEVTGSGAQACESNLQGGNGG----------GG 231
          |||:|:|               |  :  || | |          || ||          |
Db    227 RLPLEDE-------------GGEGGEGGEV-GEGG--------GGEGGPPPATSSSPPSG 264

Qy    232 VAFYNLGMNMNNFQFSGGDIYGMSGGSGGGGGGATRPAF 270
```

```
                |  ||: :    || : : |          ||||
Db    265 AGQGQLRGNMSGYDQFAGDPH-LLGWGAAAAAAPPRPAF 302


RESULT   5
Q9SR17
ID   Q9SR17      PRELIMINARY;      PRT;   315 AA.
AC   Q9SR17;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   F7O18.4 PROTEIN.
GN   F7O18.4.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,
RA   Ronning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,
RA   Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RT   "Arabidopsis thaliana chromosome III BAC F7O18 genomic sequence.";
RL   Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC011437; AAF04888.1; -.
SQ   SEQUENCE   315 AA;  32037 MW;  1766141EC657C448 CRC64;


  Query Match             45.8%;  Score 637.5;  DB 10;  Length 315;
  Best Local Similarity   50.0%;  Pred. No. 2.7e-36;
  Matches  132;  Conservative   37;  Mismatches   58;  Indels   37;  Gaps    5;

Qy     17 PTGGATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDSPNVLRSHVLEVTSGSDISE 76
            |  ||  : |       ||||||||||||||||  :|||||| |:|||:|: ||:|: |
Db     70 PREGAVEAPT-------RRPRGRPAGSKNKPKPPIFVTRDSPNALKSHVMEIASGTDVIE 122

Qy     77 AVSTYATRRGCGVCIISGTGAVTNVTIRQP------AAPAGGGVITLHGRFDILSLTGTA 130
            ::|:| ||  |:||:|| | | |||:|||      ||| |  |: | |||:||||||:
Db    123 TLATFARRRQRGICILSGNGTVANVTLRQPSTAAVAAAPGGAAVLALQGRFEILSLTGSF 182

Qy    131 LPPPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVYDRLPIEEEET 190
          || |||||: |||:|||||||||||||:| | |:|:|||:|:||:|:|| |:|||:||||
Db    183 LPGPAPPGSTGLTIYLAGGQGQVVGGSVVGPLMAAGPVMLIAATFSNATYERLPLEEEEA 242

Qy    191 PP----PRTTGVQQQQPEASQSSEVTGSGAQACESNLQGGNGGGGVAFYNLGMNMNNFQF 246
                : ||   |          | |     |   ||:|  |:  ||:  |: :
Db    243 AERGGGGGSGGVVPGQ---------LGGGGSPLSSGAGGGDGNQGLPVYNMPGNLVS--- 290

Qy    247 SGGDIYGMSGGSGGGGGGATRPAF 270
                :|||||||  : : |:
Db    291 --------NGGSGGGGQMSGQEAY 306


RESULT   6
O22130
ID   O22130      PRELIMINARY;      PRT;   317 AA.
AC   O22130;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PUTATIVE AT-HOOK DNA-BINDING PROTEIN.
GN   AT2G45430.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RX   MEDLINE=20083487; PubMed=10617197;
```

```
RA   Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA   Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA   Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA   Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA   Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA   Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA   Salzberg S.L., Fraser C.M., Venter J.C.;
RT   "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 402:761-768(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Lin X.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC002387; AAB82621.1; -.
KW   DNA-binding.
SQ   SEQUENCE   317 AA;  33519 MW;  3A3677B991AE25F0 CRC64;


  Query Match             45.2%;  Score 629.5;  DB 10;  Length 317;
  Best Local Similarity   54.1%;  Pred. No. 9.5e-36;
  Matches  126;  Conservative   37;  Mismatches   47;  Indels   23;  Gaps    5;

Qy       2 ELNRSEADEA-KAETTPTGGATSSATASGSSS-GRRPRGRPAGSKNKPKPPTIITRDSPN 59
           ::: :|   | | ::||  |  |     | :   |||||||||||||||||| |||||| |
Db      55 DIDPNEHSSAGKDQSTPGSGGESGGGGGGDNHITRRPRGRPAGSKNKPKPPIIITRDSAN 114

Qy      60 VLRSHVLEVTSGSDISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGG--VITL 117
           |:|||:|| :| |: |:|: :| ||  |:|::|| |||||||||||:  |||  |: |
Db     115 ALKSHVMEVANGCDVMESVTVFARRRQRGICVLSGNGAVTNVTIRQPASVPGGGSSVVNL 174

Qy     118 HGRFDILSLTGTALPPPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFAN 177
           ||||:||||:|: ||||||| | |||:||||||||||||:| | |:||||||:||||| |
Db     175 HGRFEILSLSGSFLPPPAPPAASGLTIYLAGGQGQVVGGSVVGPLMASGPVVIMAASFGN 234

Qy     178 AVYDRLPIEEEETPPPRTTGVQQQQPEASQSSEVTGSGAQACESNLQGGNGGG 230
           | |:|||:||::         |::|   : :: : |:             |||
Db     235 AAYERLPLEEDD---------QEEQTAGAVANNIDGNATM----------GGG 268


RESULT   7
Q9SZ70
ID   Q9SZ70      PRELIMINARY;      PRT;   339 AA.
AC   Q9SZ70;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   PUTATIVE DNA-BINDING PROTEIN.
GN   F16J13.120 OR AT4G12050.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Bevan M., Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA   Bancroft I., Mewes H.W., Mayer K.F.X., Schueller C.;
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA   Mewes H.W., Lemcke K., Mayer K.F.X.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL049638; CAB40946.1; -.
DR   EMBL; AL161533; CAB78248.1; -.
KW   DNA-binding.
SQ   SEQUENCE   339 AA;  35785 MW;  A7E3A9C73B6BBE21 CRC64;


  Query Match             44.9%;  Score 625.5;  DB 10;  Length 339;
  Best Local Similarity   49.2%;  Pred. No. 1.9e-35;
  Matches  129;  Conservative   28;  Mismatches   48;  Indels   57;  Gaps    3;

Qy       4 NRSEADEAKAETTPTGGATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDSPNVLRS 63
           | :   | |  :   |   |    ||    |||||||||||||| | |||||| | ||:
Db      88 NTNSGSEGKEMSLHGGEGGSGGGGSGEQMTRRPRGRPAGSKNKPKAPIIITRDSANALRT 147

Qy      64 HVLEVTSGSDISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGRFDI 123
           ||:|:  | || : ::|:| ||  |||::||||:||||||||| :|  | |::|||||:|
Db     148 HVMEIGDGCDIVDCMATFARRRQRGVCVMSGTGSVTNVTIRQPGSPP-GSVVSLHGRFEI 206

Qy     124 LSLTGTALPPPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVYDRL 183
           |||:|: ||||||| | ||:|||||||||||||:| | |: |||||:|||||:|| |:||
Db     207 LSLSGSFLPPPAPPAATGLSVYLAGGQGQVVGGSVVGPLLCSGPVVVMAASFSNAAYERL 266

Qy     184 PIEEEETPPPRTTGVQQQQPEASQSSEVTGSGAQACESNLQGGNGGGGVAFYNLGMNMNN 243
           |:||:|   |                                :||| ||||
Db     267 PLEEDEMQTP----------------------------VQGGGGGGG------------ 285

Qy     244 FQFSGGDIYGMSGGSGGGGGGA 265
                              |||| |:
Db     286 --------------GGGGMGS 292


RESULT   8
O82166
ID   O82166      PRELIMINARY;      PRT;   285 AA.
AC   O82166;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PUTATIVE AT-HOOK DNA-BINDING PROTEIN.
GN   AT2G35270.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RX   MEDLINE=20083487; PubMed=10617197;
RA   Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA   Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA   Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA   Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA   Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA   Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA   Salzberg S.L., Fraser C.M., Venter J.C.;
RT   "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 402:761-768(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Lin X.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC004667; AAC61811.1; -.
KW   DNA-binding.
SQ   SEQUENCE   285 AA;  29144 MW;  B7BC11E3732CB968 CRC64;


  Query Match             44.3%;  Score 616;  DB 10;  Length 285;
  Best Local Similarity   49.6%;  Pred. No. 7.1e-35;
  Matches  130;  Conservative   30;  Mismatches   64;  Indels   38;  Gaps    6;
```

```
Qy      11 AKAETTPTGGATSSATASGSSSG----RRPRGRPAGSKNKPKPPTIITRDSPNVLRSHVL 66
           |  : :  ||         |  ||    ||||||||||||||||| |:||:| | ||:|:|
Db      57 ASNDNSGLGG------GGGGGSGDLVMRRPRGRPAGSKNKPKPPVIVTRESANTLRAHIL 110

Qy      67 EVTSGSDISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGRFDILSL 126
           || || |: | :|||| ||  |:|::|||| ||||:|||| | | |:|| | |:|||| 
Db     111 EVGSGCDVFECISTYARRRQRGICVLSGTGTVTNVSIRQPT--AAGAVVTLRGTFEILSL 168

Qy     127 TGTALPPPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVYDRLPIE 186
           :|: |||||||||  ||::||| |||||||| | |:|:|||::||||| |  |:|||:: 
Db     169 SGSFLPPPAPPGATSLTIFLAGAQGQVVGGNVVGELMAAGPVMVMAASFTNVAYERLPLD 228

Qy     187 EEETPPPRTTGVQQQQPEASQSSEVTGSGAQACESNLQGGNGGGGVAFYNLGMNMNNFQF 246
           | |              |  ||    | |    |:      |||| |:|| |:|
Db     229 EHE-------------EHLQSGGGGGGGNMYSEAT----GGGGGLPFFNLPMSMPQI-- 268

Qy     247 SGGDIYGMSGGSGGGGGGATRP 268
                |:    |  |    |
Db     269 ------GVESWQGNHAGAGRAP 284


RESULT  9
O49662
ID   O49662      PRELIMINARY;      PRT;   324 AA.
AC   O49662;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PUTATIVE DNA BINDING PROTEIN.
GN   T12H17.200 OR AT4G22810.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Bevan M., Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA   Bancroft I., Mewes H.W., Mayer K., Schueller C.;
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA   Mewes H.W., Lemcke K., Mayer K.F.X.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL021635; CAA16566.1; -.
DR   EMBL; AL161558; CAB79236.1; -.
SQ   SEQUENCE   324 AA;  34339 MW;  B8EF7520B474C528 CRC64;


  Query Match             43.5%;  Score 605;  DB 10;  Length 324;
  Best Local Similarity   56.8%;  Pred. No. 4.6e-34;
  Matches  129;  Conservative   26;  Mismatches   42;  Indels   30;  Gaps    6;

Qy      19 GGATSSATASGSSSG-----RRPRGRPAGSKNKPKPPTIITRDSPNVLRSHVLEVTSGSD 73
           ||   |    | | |     |||||||||||||||||| |||||| | ||:||:|:  | |
Db      88 GG---SGEGGGGSGGDHQMTRRPRGRPAGSKNKPKPPIIITRDSANALRTHVMEIGDGCD 144

Qy      74 ISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAA-PAGGGVITLHGRFDILSLTGTALP 132
           : |:|:|:| ||  |||::|||| ||||||||| : |: | |::|||||:||||:|: ||
Db     145 LVESVATFARRRQRGVCVMSGTGNVTNVTIRQPGSHPSPGSVVSLHGRFEILSLSGSFLP 204

Qy     133 PPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVYDRLPIEEEE--T 190
           ||||| | ||:|||||||||||||:| | |: :||||:|||||:|| |:|||:||:|  |
Db     205 PPAPPTATGLSVYLAGGQGQVVGGSVVGPLLCAGPVVVMAASFSNAAYERLPLEEDEMQT 264

Qy     191 P------------PPRTTGVQQQQPEASQSSEVTGSGAQACESNLQG 225
           |            ||      | | :|        || |    || |
Db     265 PVHGGGGGGSLESPPMMGQQLQHQQQAM-------SGHQGLPPNLLG 304
```

```
RESULT  10
Q9LZX7
ID   Q9LZX7      PRELIMINARY;      PRT;   265 AA.
AC   Q9LZX7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   HYPOTHETICAL 28.4 KDA PROTEIN.
GN   T4C21_280.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Choisne N., Robert C., Brottier P., Wincker P., Cattolico L.,
RA   Artiguenave F., Saurin W., Weissenbach J., Mewes H.W., Rudd S.,
RA   Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL162295; CAB82691.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   265 AA;  28412 MW;  31162D06E4BAD0AD CRC64;


  Query Match             41.7%;  Score 581;  DB 10;  Length 265;
  Best Local Similarity   50.2%;  Pred. No. 1.6e-32;
  Matches  119;  Conservative   31;  Mismatches   57;  Indels   30;  Gaps    3;

Qy      34 RRPRGRPAGSKNKPKPPTIITRDSPNVLRSHVLEVTSGSDISEAVSTYATRRGCGVCIIS 93
           |||||||||||||| | |:|||| |  | ||:|:|:  |: |::: :| ||  |||::: 
Db      59 RRPRGRPAGSKNKPKAPIIVTRDSANAFRCHVMEITNACDVMESLAVFARRRQRGVCVLT 118

Qy      94 GTGAVTNVTIRQPAAPAGGGVITLHGRFDILSLTGTALPPPAPPGAGGLTVYLAGGQGQV 153
           | |||||||:|||    ||||::|||||:||||:|: ||||||| | || ||||||||||
Db     119 GNGAVTNVTVRQP----GGGVVSLHGRFEILSLSGSFLPPPAPPAASGLKVYLAGGQGQV 174

Qy     154 VGGNVAGSLIASGPVVLMAASFANAVYDRLPIEEEETPPPRTTGVQQQQPEASQSSEVTG 213
           :||:| | | || |||:||||| || |:|||:|||| |: |
Db     175 IGGSVVGPLTASSPVVVMAASFGNASYERLPLEEEE----------------ETEREIDG 218

Qy     214 SGAQACESNLQGGNGGGGVAFYNLGMNMNNFQFSGGDIYGMSGGSGGGGGGATRPAF 270
           : |:|  :  |        :|     |: |     |: |           |  ||:|
Db     219 NAARAIGTQTQKQLMQDATSFIGSPSNLINSVSLPGEAY----------WGTQRPSF 265


RESULT  11
Q9M2S3
ID   Q9M2S3      PRELIMINARY;      PRT;   310 AA.
AC   Q9M2S3;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   HYPOTHETICAL 32.6 KDA PROTEIN.
GN   T22E16.220.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Benes V., Wurmbach E., Drzonek H., Ansorge W., Mewes H.W., Lemcke K.,
RA   Mayer K.F.X., Quetier F., Salanoubat M.;
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
```

```
RA   EU Arabidopsis sequencing project;
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL132975; CAB75914.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   310 AA;  32603 MW;  0F856170E37A40CB CRC64;
```

Query Match 41.1%; Score 572; DB 10; Length 310;
Best Local Similarity 47.8%; Pred. No. 8e-32;
Matches 120; Conservative 44; Mismatches 69; Indels 18; Gaps 6;

```
Qy       5 RSEADEAKAETTPTGGATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDSPNVLRSH 64
Db      63 QSQEEQNSRDEQP---AVEPGSGSG-STGRRPRGRPPGSKNKPKSPVVVTKESPNSLQSH 118

Qy      65 VLEVTSGSDISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGRFDIL 124
Db     119 VLEIATGADVAESLNAFARRRGRGVSVLSGSGLVTNVTLRQPA--ASGGVVSLRGQFEIL 176

Qy     125 SLTGTALPPP-APPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVYDRL 183
Db     177 SMCGAFLPTSGSPAAAAGLTIYLAGAQGQVVGGGVAGPLIASGPVIVIAATFCNATYERL 236

Qy     184 PIEEEETPPPRTTGVQQQQPEASQSSEVTGSGAQACESNLQGGNGGGGVAFYNLGMNM-- 241
Db     237 PIEEEQ---------QQEQPLQLEDGKKQKEENDDNESGNNGNEGSMQPPMYNMPPNFIP 287

Qy     242 NNFQFSGGDIY 252
Db     288 NGHQMAQHDVY 298
```

```
RESULT  12
Q9SJG4
ID   Q9SJG4      PRELIMINARY;      PRT;   257 AA.
AC   Q9SJG4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   AT2G42940 PROTEIN.
GN   AT2G42940.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RX   MEDLINE=20083487; PubMed=10617197;
RA   Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA   Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA   Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA   Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA   Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA   Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA   Salzberg S.L., Fraser C.M., Venter J.C.;
RT   "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 402:761-768(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Lin X.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC006931; AAD21715.1; -.
DR   InterPro; IPR000637; AT_hook.
DR   Pfam; PF02178; AT_hook; 1.
DR   SMART; SM00384; AT_hook; 1.
SQ   SEQUENCE   257 AA;  27004 MW;  00DA8F50213F6DE9 CRC64;
```

Query Match 39.8%; Score 554; DB 10; Length 257;
Best Local Similarity 52.8%; Pred. No. 1.1e-30;
Matches 113; Conservative 30; Mismatches 37; Indels 34; Gaps 4;

```
Qy      16 TPTGGATSS-------ATASGSSS--------------------GRRPRGRPAGSKNKPK 48
Db       8 TPTSVGSKSVPMRNHEATERGNTNNNLRALPKAVQPVSSIEGEMAKRPRGRPAGSKNKPK 67

Qy      49 PPTIITRDSPNVLRSHVLEVTSGSDISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAA 108
Db      68 PPIIVTHDSPNSLRANAVEISSGCDICETLSDFARRKQRGLCILSANGCVTNVTLRQPA- 126

Qy     109 PAGGGVITLHGRFDILSLTGTALPPPAPPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPV 168
Db     127 -SSGAIVTLHGRYEILSLLGSILPPPAPLGITGLTIYLAGPQGQVVGGGVVGGLIASGPV 185

Qy     169 VLMAASFANAVYDRLPIEEEETPPPRTTGVQQQQ 202
Db     186 VLMAASFMNAVFDRLPMDDDE-----AASMQNQQ 214
```

```
RESULT  13
Q9LTA2
ID   Q9LTA2      PRELIMINARY;      PRT;   276 AA.
AC   Q9LTA2;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   SIMILARITY TO AT-HOOK DNA-BINDING PROTEIN.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COLUMBIA;
RA   Kaneko T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,
RA   Tabata S.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 5. XI.";
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB025613; BAA98158.1; -.
KW   DNA-binding.
SQ   SEQUENCE   276 AA;  29390 MW;  A804D55ADA0568BF CRC64;
```

Query Match 35.9%; Score 499.5; DB 10; Length 276;
Best Local Similarity 43.8%; Pred. No. 6.5e-27;
Matches 106; Conservative 41; Mismatches 66; Indels 29; Gaps 5;

```
Qy      21 ATSSATASGSSSG--RRPRGRPAGSKNKPKPPTIITRDSPNVLRSHVLEVTSGSDISEAV 78
Db      41 STVTPTVDDSSIEVVRRPRGRPPGSKNKPKPPVFVTRDTDPPMSPYILEVPSGNDVVEAI 100

Qy      79 STYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGRFDILSLTGTALPPP---- 134
Db     101 NRFCRRKSIGVCVLSGSGSVANVTLRQPSPAALGSTITFHGKFDLLSVSATFLPPPPRTS 160

Qy     135 -APPGAGGLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVYDRLPIEEEETPPP 193
Db     161 LSPPVSNFFTVSLAGPQGQIIGGFVAGPLISAGTVYVIAASFNNPSYHRLPAEEEQKHSA 220

Qy     194 RTTGVQQQQPEAS----QSSEVTGSGAQACESNLQGGNGGGGVAFYNLGMNMNNFQFSGG 249
Db     221 GTGEREGQSPPVSGGGEESGQMAGSGGESC-----------GVSMYSCHM-------GGS 262

Qy     250 DI 251
Db     263 DV 264
```

```
RESULT  14
Q9M9R4
ID   Q9M9R4      PRELIMINARY;      PRT;   206 AA.
```

```
AC   Q9M9R4;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   F14L17.27 PROTEIN.
GN   F14L17.27.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Liu S., Sakano H., Yu G., Lee J., Lenz C., Pham P., Toriumi M.,
RA   Chin C., Chiou J., Choi E., Chung M., Gonzalez A., Howng B., Liu A.,
RA   Vaysberg M., Altafi H., Brooks S., Buehler E., Chao Q., Conn L.,
RA   Conway A., Hansen N., Johnson-Hopson C., Khan S., Kim C., Lam B.,
RA   Miranda M., Nguyen M., Palm C., Shinn P., Southwick A., Davis R.,
RA   Ecker J., Federspiel N., Theologis A.;
RT   "The sequence of BAC F14L17 from Arabidopsis thaliana chromosome 1.";
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Theologis A.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC012188; AAF43950.1; -.
DR   InterPro; IPR000637; AT_hook.
DR   SMART; SM00384; AT_hook; 1.
SQ   SEQUENCE   206 AA;  21459 MW;  CE20976A467287C8 CRC64;


  Query Match             30.0%;  Score 417;  DB 10;  Length 206;
  Best Local Similarity   46.0%;  Pred. No. 2.1e-21;
  Matches   91;  Conservative   32;  Mismatches   59;  Indels   16;  Gaps    5;

Qy      35 RPRGRPAGSKNKPKPPTIITRDSPNVLRSHVLEVTSGSDISEAVSTYATRRGCGVCIISG 94
           |||||| ||||||| |  :| | |  :  ::||| ||:|: ||:: :   :  | |::||
Db       6 RPRGRPRGSKNKPKAPIFVTIDPP--MSPYILEVPSGNDVVEALNRFCRGKAIGFCVLSG 63

Qy      95 TGAVTNVTIRQPAAPAGGGVITLHGRFDILSLTGTALPP-----PAPPGAGGLTVYLAGG 149
           :|:| :||:|||:  | |  || ||:||:||:: | |||      :|| :   || |||
Db      64 SGSVADVTLRQPSPAAPGSTITFHGKFDLLSVSATFLPPLPPTSLSPPVSNFFTVSLAGP 123

Qy     150 QGQVVGGNVAGSLIASGPVVLMAASFANAVYDRLPIEEEETPPPRTTGVQQQQPEASQSS 209
           ||:|:|| ||| |:|:| |  :| || |  | |||  |||    | :    :  | ||
Db     124 QGKVIGGFVAGPLVAAGTVYFVATSFKNPSYHRLPATEEE---QRNSA---EGEEEGQSP 177

Qy     210 EVTGSGAQACESNLQGGN 227
           |:| |    ||   ||:
Db     178 PVSGGGG---ESMYVGGS 192


RESULT  15
Q941Y9
ID   Q941Y9      PRELIMINARY;      PRT;   265 AA.
AC   Q941Y9;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   P0431G06.21 PROTEIN.
GN   P0431G06.21.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. NIPPONBARE;
RA   Sasaki T., Matsumoto T., Yamamoto K.;
RT   "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT   clone:P0431G06.";
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP003683; BAB64709.1; -.
SQ   SEQUENCE   265 AA;  26638 MW;  1BB4ED3E30AD6CAA CRC64;


  Query Match             29.2%;  Score 407;  DB 10;  Length 265;
  Best Local Similarity   36.8%;  Pred. No. 1.3e-20;
  Matches   96;  Conservative   40;  Mismatches   83;  Indels   42;  Gaps    5;

Qy       7 EADEAKAETTPTGGATSSATASGSSSGRRPRGRPAGSKNKPKPPTIITRDSPNVLRSHVL 66
           ||  | |   |             '|  |:||||| |||||||||| ::||:|   :|  ||
Db      13 EASPAPALDLP------------SPPRKPRGRPLGSKNKPKPPVVVTRESEAAMRPVVL 59

Qy      67 EVTSGSDISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGRFDILSL 126
           |: :| ::: ||: :| ||  || :: | | |  ||:| | :|     : |||||::|||
Db      60 ELGAGCEVAAAVAAFARRRRVGVSVLCGRGTVAAVTLRLPTSPP--AAVKLHGRFEVLSL 117

Qy     127 TGTALPPPAPPGAG---GLTVYLAGGQGQVVGGNVAGSLIASGPVVLMAASFANAVYDRL 183
           :|| ||  |  ||       :| |||  |||:|| :|| :  :  :|:::||:| :|   ||
Db     118 SGTVLPSAAGEGAAPPPPFSVSLAGAGGQVIGGTLAGEMTTADGLVVVAATFGSAEVHRL 177

Qy     184 PIEEEETPPPRTTGVQQQQPEASQSSEVTGSGAQACESNLQGGNGGGGVAFYNLGMNMNN 243
           | :|::       | ::: |:  |    | :   | :  | |  || |||
Db     178 PADEDDEATGSRGGEERRHPQ--QQPPQTVAATSAVDVGLLGYGGGVGVA---------- 225

Qy     244 FQFSGGDIYGMSGGSGGGGGG 264
                     ||: ||  |
Db     226 -----------GGASGGQVG 234


Search completed: July 10, 2002, 22:28:40
Job time: 572 sec
```